Journal of the
OPTICAL SOCIETY
of AMERICA

VOLUME 39, NUMBER 10 OCTOBER, 1949

The Electro-Optic Effect in Uniaxial Crystals of the Type X H_2PO_4. I. Theoretical

BRUCE H. BILLINGS
Research Laboratory, Baird Associates, Inc., Cambridge 38, Massachusetts
(Received June 22, 1949)

When voltage is put across a crystal of the type X H_2PO_4, the index ellipsoid in the crystal is rotated and the length of the axes changes. Two electro-optic coefficients are necessary to describe the effect in the most general plate of the crystal. The necessity for two coefficients is shown from consideration of the crystal symmetry. The behavior of Z-cut and X-cut plates are treated in detail. For Z-cut plates with the voltage in the *Z* direction the crystal becomes biaxial with the plane of the optic axes at 45° to the *X* and *Y* crystallographic axis. The retardation along the normal to the plate is directly proportional to the voltage and independent of the thickness. The characteristics of the polarization interference pattern for excited X and Z plates are discussed.

INTRODUCTION

THERE are many places where an electric light valve or shutter has useful applications. One of the most obvious forms of light valve is the Kerr cell. This cell contains a liquid between a pair of parallel plate electrodes. When an electric field is applied, the liquid becomes birefringent. If the cell is placed between crossed polarizers, light will not pass through the combination until voltage is applied. There are many difficulties associated with the ordinary liquid Kerr cell. This liquid cell uses nitrobenzene which has a high Kerr constant only when it is extremely pure. The operation of the cell tends to decompose the material and the most successful cells have been those in which the nitrobenzene is continuously redistilled as the cell is used. In addition, the liquid has a disagreeable odor and is intensely poisonous.

A solid electro-optic shutter avoids many of these difficulties. As a result, such a device becomes a practical solution for many different problems and should be a useful scientific tool.

The Kerr effect in most solids is extremely small. In certain crystals, however, there is a large electro-optic effect which requires voltage no higher than those required in the liquid cell. This effect varies with first power of the electric field rather than with the square of the field as does the normal Kerr effect.

During the last few years several papers[1,2,3,4] have appeared discussing the electro-optic effect in tetragonal scalenohedral crystals of the type X H_2PO_4. Now that some of these crystals such as $NH_4H_2PO_4$ or ADP[5] are available as light valves, it seems appropriate to give a more complete treatment of the useful effect than is available in any of these papers.

THEORETICAL

The birefringence of a crystal can be described in terms of its index ellipsoid. The index ellipsoid has axes which are equal to the three crystal indices. If a plane is passed through the center of this ellipsoid at right angles to the direction in which a ray of light is going, it will cut the ellipsoid in an ellipse whose major and minor axes are the two indices associated with that particular direction in the crystal. In the case of a biaxial crystal the equation for this ellipsoid can be written

$$a^2x^2+b^2y^2+c^2z^2=1 \qquad (1)$$

where a, b and c are the three reciprocal indices of

[1] G. D. Gottschall, J. Soc. Mot. Pict. Eng. **51**, 13 (1948).
[2] Hans Jaffe, Phys. Rev. **73**, 95 (1948).
[3] B. Zwicker and P. Scherrer, Helv. Phys. Acta **16**, 214 (1943).
[4] B. Zwicker and P. Scherrer, Helv. Phys. Acta **17**, 346 (1944).
[5] This crystal is grown under the trademark "PN" by the Brush Development Company, 3405 Perkins Avenue, Cleveland, Ohio.

refraction. In a uniaxial crystal two of the coefficients become equal and the equation reduces to

$$o^2(x^2+y^2)+e^2z^2=1 \tag{2}$$

where o is the ordinary reciprocal index and e the extraordinary reciprocal index. The electro-optic effect in crystals is a change in crystalline birefringence which occurs when a crystal is placed in an electric field. The first general treatment of the effect was given by Pockels.[6,7] He showed the number of constants to be expected in the different crystal classes and made experimental measurements in the case of quartz, $NaClO_3$ and rochelle salts. This change can be described in terms of a change of the orientation and the dimensions of the index ellipsoid. The numerical constants involved in this change can be considered as small corrections to be applied to the index ellipsoid constants. In order to write them it is necessary to put the equation of the index ellipsoid in its most general form. Crystallographers commonly refer the faces of a crystal to a cartesian coordinate system. The directions of these crystallographic axes are prescribed for each crystal class. Instead of referring the ellipsoid to its own principal axes it is referred instead to these crystallographic symmetry axes. The equation of the ellipsoid then becomes

$$1=a_{11}x^2+a_{22}y^2+a_{33}z^2+2a_{23}yz+2a_{31}zx+2a_{12}xy. \tag{3}$$

The new constants, a_{11}, a_{12}, etc., are called the polarization constants. If the a's are all known, it is possible to find the orientation of the axes of the ellipsoid with respect to the crystallographic axes.

The direction cosines connecting the two sets of axes may be described in the matrix

	X'	Y'	Z'
X	α_1	α_2	α_3
Y	β_1	β_2	β_3
Z	γ_1	γ_2	γ_3

(4)

where the primed axes are the crystallographic axes, the unprimed axes are the axes of the index ellipsoid, and α_1 is the cosine of the angle between x and x', etc. When an electric field is applied, the polarization constants can be written

$$\begin{aligned}
a_{11}-a_0^2&=r_{11}E_x+r_{12}E_y+r_{13}E_z\\
a_{22}-b_0^2&=r_{21}E_x+r_{22}E_y+r_{23}E_z\\
a_{33}-c_0^2&=r_{31}E_x+r_{32}E_y+r_{33}E_z\\
a_{23}&=r_{41}E_x+r_{42}E_y+r_{43}E_z\\
a_{31}&=r_{51}E_x+r_{52}E_y+r_{53}E_z\\
a_{12}&=r_{61}E_x+r_{62}E_y+r_{63}E_z
\end{aligned} \tag{5}$$

where the r_{ij} are the electro-optic coefficients and E_x, E_y, E_z are the components of the field along the three crystallographic axes. These equations have been written for a crystal in which the crystallographic axes coincide with the polarization axes. They differ from the equations of Pockels by referring to the field instead of the electrical polarization. The constants r_{ij} are thus related to Pockels' constants e_{ij} by the equation

$$r_{ij}=\frac{K_j-1}{4\pi}e_{ij}$$

where K_j is the dielectric constant in the direction in which the field is applied. The symbol r_{ij} is used in accordance with Cady's[8] suggestion that the Mueller[9] convention be used. If no field were applied the constants would reduce to the three coefficients given in Eq. (1).

The most striking characteristic of a crystal is its symmetry. This symmetry is described by a set of operations sometimes called "covering operations" which can be performed on the crystal without altering its appearance or physical behavior. For example, if a cube is rotated 90° about an axis normal to any face the cube will not appear to have been moved. If the cube is isotropic with respect to all its other physical properties, the operations which leave its external appearance unchanged will be the characteristic symmetry operations. If the cube is made of wood which has a grain perpendicular to one face, the tensile strength would not be the same in the case of the 90° rotation about an axis parallel to this face. In general the physical characteristics of a crystal possess the same or higher symmetry as the symmetry of its form. This is,

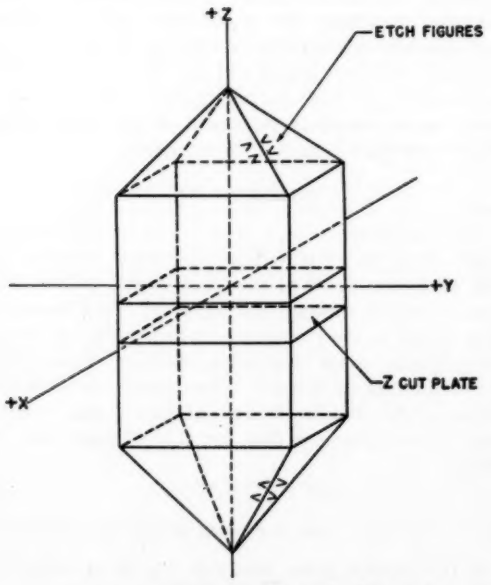

FIG. 1. Sketch of an $NH_4H_2PO_4$ crystal with its crystallographic axes labelled.

[6] F. Pockels, Abhandlungen der Gesellschaft der Wissenschaften zu Gottingen, 39, 1 (1893).
[7] F. Pockels, *Lehrbuch der Kristalloptik* (Leipzig, 1906).
[8] W. G. Cady, *Piezoelectricity* (McGraw-Hill Book Company, Inc., New York, 1946).
[9] H. Mueller, Phys. Rev. 58, 805–811 (1940).

of course, true only for an ideal crystal of any material.

Since many readers are perhaps rusty in their crystallography, it seems apropos to carry through the actual reduction of the constants in Eq. (4). In Fig. 1 is a sketch of an $NH_4H_2PO_4$ crystal with its crystallographic axes labeled. At first glance the crystal appears to possess a center of symmetry, i.e., it looks as though every point in the crystal could be reflected through the center to an equal distance on the other side and the resulting crystal would be identical. This turns out not to be the case. It is most simply demonstrated by etch figures. If one of the end faces is slightly moistened and dried, a series of little pits appear. These pits are triangular in shape and are oppositely oriented on opposite faces. There are two main symmetry operations which can be performed on the crystal. The first is a rotation through $\pi/2$ about the z axis followed by an inversion through the center. This, of course, does not mean that the crystal has a center of symmetry because the inversion must always be preceded by the $\pi/2$ rotation about z. The second operation is a rotation of π about either the x or y axis. Crystals whose characteristics are unchanged after these operations are said to belong in the class V_d or $42m$. If these operations are now performed on Eq. (5) the constants must remain unchanged. The operation of rotating through an angle π about the x axis leaves x unchanged and changes y to $-y$ and z to $-z$. The field E_x is unchanged whereas E_y becomes $-E_y$ and E_z becomes $-E_z$. When these substitutions are made in Eq. (5) and Eq. (3) the coefficients a_{11} through a_{12} must change sign in such a way that Eq. (3) is unchanged. The effect on the electro-optic coefficients may be simply shown by writing Eq. (5) in matrix form. For further aid the multiplier of the a_{ij}'s in Eq. (3) is placed in the a_{ij} vector.

$$\begin{vmatrix} a_{11}(x^2) \\ a_{22}(y^2) \\ a_{33}(z^2) \\ a_{23}(yz) \\ a_{31}(xz) \\ a_{12}(xy) \end{vmatrix} = \begin{vmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \\ r_{41} & r_{42} & r_{43} \\ r_{51} & r_{52} & r_{53} \\ r_{61} & r_{62} & r_{63} \end{vmatrix} \cdot \begin{vmatrix} E_x(x) \\ E_y(y) \\ E_z(z) \end{vmatrix}.$$

If the substitutions $x \rightarrow x$, $y \rightarrow -y$, $z \rightarrow -z$ are made in this matrix equation, it is apparent that a_{11}, a_{22} and a_{33} will not change signs. Since E_y and E_z do change sign, this means that all the r_{ij}'s in the first three rows are zero with the exception of r_{11}, r_{21}, and r_{31}. The operation of rotating the crystal through an angle π about the y axis changes the sign of x and z but leaves y unchanged. This operation makes these the remaining three coefficients in the first three rows vanish. The final array of elements in the matrix becomes

$$\begin{vmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ r_{41} & 0 & 0 \\ 0 & r_{41} & 0 \\ 0 & 0 & r_{63} \end{vmatrix}. \tag{7}$$

The constants a_{ij} thus reduce to

$$\begin{aligned} a_{11}=a_{22}=o^2 \quad & a_{23}=r_{41}E_x \\ a_{33}=e^2 \quad & a_{31}=r_{41}E_y \\ & a_{12}=r_{63}E_z. \end{aligned} \tag{8}$$

FIELD IN Z DIRECTION

The crystal section which is most useful as a shutter is the Z-cut or basal plate with field applied in the Z direction and the light traveling in the same direction. For light thus parallel to the optic axis the plate has no retardation and appears isotropic. The analysis below shows that in this direction the electrically induced retardation is also highest for a given voltage.

For this case the fields E_x and E_y become zero and the only non-zero coefficients are thus a_{11}, a_{22}, a_{33} and a_{12}. These values can be inserted into Eq. (3) to give

$$e^2z^2+o^2(x^2+y^2)+2r_{63}E_zxy=1. \tag{9}$$

From this equation it is apparent that the Z and Z' axes coincide. This shows that $\gamma_1=\gamma_2=\alpha_3=\beta_3=0$ and $\gamma_3=1$.

To find the remaining cosines and hence the angle by which the ellipsoid is rotated, Eq. (9) is transformed to the coordinate system which has its axes along the ellipsoid axes. This equation connecting the two sets can be written down directly from the matrix (4) which represents the direction cosines between them.

$$\begin{aligned} x' &= \alpha_1x+\beta_1y \\ y' &= \alpha_2x+\beta_2y. \end{aligned}$$

Since the rotation is in a plane

$$-\alpha_2=\beta_1; \quad \alpha_1=\beta_2.$$

These equations can now be put in Eq. (9) and the cross terms set equal to zero so that the axes will coincide with the ellipsoid axes.

$$2r_{63}C(\alpha_1^2-\alpha_2^2)xy=0. \tag{10}$$

The expression for the angle then becomes

$$\cos 2\alpha=0 \quad \alpha=\pm 45°. \tag{11}$$

Since this rotation is independent of the field it is possible at this stage of the analysis to say what will be the appearance between polarizers of a z-cut slab of PN when a field is applied in the z direction. The crystal becomes biaxial and the characteristic uniaxial fringe pattern of circles is replaced by a series of ovals. The long axis of the oval lies at 45° to the crystallographic axis. The length of the long axes of the ovals in the pattern will change as the field is changed. The direction, however, remains constant. To find how the axis length changes with the field it is necessary to compute the actual value of the crystal indices.

It is necessary now to find the new reciprocal indices, a, b and c. These are related to the polarization constants by formulas which can be written directly

from Eq. (3) and the matrix (4)

$$\begin{aligned} a^2&=a_{11}\alpha_1^2+a_{22}\alpha_2^2+a_{33}\alpha_3^2+2a_{23}\alpha_2\alpha_3 \\ &\quad +2a_{31}\alpha_3\alpha_1+2a_{12}\alpha_1\alpha_2 \\ b^2&=a_{11}\beta_1^2+a_{22}\beta_2^2+a_{33}\beta_3^2+2a_{23}\beta_2\beta_3 \\ &\quad +2a_{31}\beta_3\beta_1+2a_{12}\beta_1\beta_2 \\ c^2&=a_{11}\gamma_1^2+a_{22}\gamma_2^2+a_{33}\gamma_3^2+2a_{23}\gamma_2\gamma_3 \\ &\quad +2a_{31}\gamma_3\gamma_1+2a_{12}\gamma_1\gamma_2. \end{aligned} \tag{12}$$

Since the only rotation is ϕ_z in the xy plane,

$$\begin{aligned} a^2&=a_{11}\cos^2\phi+a_{22}\sin^2\phi+2a_{12}\sin\phi\cos\phi \\ &=o^2\cos^2\phi+o^2\sin^2\phi+2r_{63}E_z\sin\phi\cos\phi \\ b^2&=a_{11}\sin^2\phi+a_{22}\cos^2\phi-2a_{12}\sin\phi\cos\phi \\ &=o^2\sin^2\phi+o^2\cos^2\phi-2r_{63}E_z\sin\phi\cos\phi \\ c^2&=a_{33}=e^2 \end{aligned} \tag{13}$$

since $\phi_z=-45°$

$$a^2=o^2-2r_{63}E_z\frac{1}{\sqrt{2}}\cdot\frac{1}{\sqrt{2}}=o^2-r_{63}E_z$$

$$b^2=o^2+2r_{63}E_z\frac{1}{\sqrt{2}}\cdot\frac{1}{\sqrt{2}}=o^2+r_{63}E_z. \tag{14}$$

One can write

$$\begin{aligned} c^2&=e^2=\frac{1}{\epsilon^2}=\frac{1}{\alpha^2} \\ a^2&=\frac{1}{\gamma^2}\quad b^2=\frac{1}{\beta^2}. \end{aligned} \tag{15}$$

Since the only field is in the Z direction, E_z can be replaced by V/d when V is the applied voltage and d is the plate thickness. The value of r_{63} then includes the dielectric constant.

The retardation which will appear along the normal to the basal section after the voltage is applied can be written:

$$\Gamma=\frac{d}{\lambda}\left(\frac{1}{a}-\frac{1}{b}\right). \tag{16}$$

This can be rewritten:

$$\begin{aligned} \Gamma&=\frac{d}{\lambda}\frac{a^2-b^2}{(a+b)ab} \\ &=\frac{d}{2\lambda}(a^2-b^2)\omega^3. \end{aligned} \tag{17}$$

If Eq. (13) is inserted in this final equation, one gets the equation:

$$\Gamma=\frac{r_{63}V\omega^3}{\lambda}=pV. \tag{18}$$

This shows that in this particular cut of crystal the retardation along the normal is a function of the voltage rather than the field strength. In other words, the retardation for a given voltage is independent of the thickness of the plate.

The value for r_{63} becomes:

$$r_{63}=\lambda\Gamma/\omega^3V \tag{19}$$

The value of this constant has been determined for $NH_4H_2PO_4$ or "PN" under a variety of conditions by Dr. Hans Jaffe of The Brush Development Company.

The value of the new indices can now be calculated from Eq. (13):

$$\gamma=\frac{1}{a}=\frac{1}{\left(o^2-\dfrac{\lambda\Gamma}{\omega^3d}\right)^{\frac{1}{2}}} \tag{20}$$

or

$$\gamma=\omega+\frac{\lambda\Gamma}{2d}. \tag{21}$$

Similarly,

$$\beta=\omega-\frac{\lambda\Gamma}{2d}. \tag{22}$$

One of the striking things about these two index values is the small change which occurs with the application of voltage. A half-wave retardation requires nearly 10,000 volts for PN and this retardation is the result of index changes of only a few parts in a million for plate thicknesses of the order of a centimeter.

Although this value seems small, it is larger in ADP than in other crystals which have been previously studied. The coefficient r_{63} in ADP is 8.2×10^{-7} whereas the largest coefficients for quartz and rochelle salts are $r_{11(\text{quartz})}=4.67\times10^{-8}$ cm/kv and $r_{52(\text{Rochelle})}=-36.0\times10^{-8}$ cm/kv.

When voltage is applied to the crystal it becomes biaxial with an axial angle Ω. The retardation along the normal which was expressed in Eq. (14) can be written in terms of this axial angle. The expression for the retardation as a function of angle of incidence in a Z-cut slab of biaxial crystal can be written:

$$\frac{\Gamma}{kd}=\sin\theta_1\sin\theta_2/\cos r \tag{23}$$

where θ_1 and θ_2 are the angles made by the ray with the two axes; r is the angle of incidence of the ray; d is the thickness of the slab, k is a constant. When $r=0$, this equation becomes:

$$\Gamma=kd\sin^2\Omega \tag{24}$$

where Ω is the angle between the optic axes and the Z axis. In combination with Eq. (14) this becomes:

$$kd\sin^2\Omega=pV \tag{25}$$

where p is a constant and V is the voltage across the crystal plate.

It is apparent from these equations that the value of the axial angle is inversely proportional to the square

root of the thickness and directly proportional to the square root of the voltage.

FIELD IN *X* OR *Y* DIRECTIONS

Since the equations are symmetrical, the effect produced by a field in the X direction is the same as a field in the Y direction. Accordingly, this section will be limited to a discussion of the effect of a field along X. In this case the only non-zero coefficients are a_{11}, a_{22}, a_{32} and a_{23}. These values can be inserted into Eq. (3) to give

$$o^2(x^2+y^2)+e^2z^2+2r_{41}E_xyz=1. \tag{26}$$

From this equation it is apparent that the X and X' axes coincide and the only rotation is in the yz plane. From an analysis similar to that used in the case of the basal section it can be shown that the axes are rotated through an angle

$$\tan 2\xi=-\frac{2r_{41}E_x}{e^2-o^2}. \tag{27}$$

This differs from the previous case in that the angle is dependent on the field and that it is quite small. Its size can, of course, only be stated with a knowledge of r_{41}. This coefficient will be discussed in a later paper. For $NH_4H_2PO_4$ it is of such a size that the rotation is only a few minutes.

The reciprocal indices can be found from Eq. (11) which can be rewritten

$$\begin{aligned} b^2&=o^2\\ a^2&=o^2\cos^2\xi+e^2\sin^2\xi+2r_{41}E_x\cos\xi\sin\xi\\ c^2&=o^2\sin^2\xi+e^2\cos\xi-2r_{41}E_x\cos\xi\sin\xi. \end{aligned} \tag{28}$$

As in the first case this represents a biaxial crystal. To go from Eq. (12) to Eq. (28) it is necessary to interchange a and b. This is because b is by definition intermediate between a and c. If the field had been applied in the Y direction, no such interchange would have been necessary. Since the axial plane is always at right angles to b and since the direction of b is independent of the field, the plane of the axis is again independent of the field as in the Z case. By inspection of Eq. (26) it is apparent that the section of the ellipsoid perpendicular to Z is a circle, and the position of this circular section is independent of the field. This indicates that one of the optic axes of the crystal coincides with the original optic axis and is independent of the field. The other axis moves from this first axis as the field is applied. The axial angle which is the angle between the ellipsoid axis and the optic axis is exactly ξ. This can also be checked by using the equation

$$\sin^2\Omega=b^2-c^2/a^2-c^2. \tag{29}$$

Observation in the Z direction thus does not give a useful effect when field is applied in the X direction. Furthermore, the axial angle of the new biaxial crystal is only a few minutes as opposed to several degrees which is obtained when the field is applied in the Z direction.

In the Z-cut case the retardation was zero in the Z direction until the field was applied in this direction. In the X-cut case the retardation in the X direction is already large. For practical use the change in retardation for a reasonable voltage must be large. The retardation without field is:

$$\Gamma=(\epsilon-\omega)d/\lambda. \tag{30}$$

The change in retardation may be calculated to be

$$\Delta\Gamma=\tfrac{1}{2}r_{41}{}^2\frac{V^2}{d}\frac{\omega^2\epsilon^2(\omega^3+\epsilon^3)}{\omega^2-\epsilon^2}. \tag{31}$$

This turns out to be of the order of 10^{-3} for a PN crystal 1-cm thick with 25 kv applied in the x direction. Even in an isotropic medium such low induced birefringence would not be particularly useful.

CONCLUSIONS

The conclusions of these analyses can be summarized as follows:

For the field in the Z direction:

1. Uniaxial crystals of the form X H_2PO_4 become biaxial.
2. The plane of the optic axes is independent of the magnitude of the voltage and is inclined at 45° to the crystallographic axis.
3. For rays parallel to the field and at a given wavelength the retardation through the plate depends only on the voltage and is independent of the plate thickness.

For a field in the X direction:

4. The crystal again becomes biaxial.
5. The plane of the axes is independent of the magnitude of the voltage and is perpendicular to the X axis.
6. The retardation change through the plate for rays parallel to the field depends on the square of the applied voltage and the inverse first power of the plate thickness.
7. If the r_{41} and r_{63} coefficients are of the same order of magnitude a 1-cm plate of ADP gives a retardation change of the order of 10^{-3} for 25 kv applied voltage.

The problem of measuring the coefficients and of determining experimentally the properties of the specific crystal ADP will be discussed in the next paper.

In conclusion the author wishes to express appreciation to Miss Elizabeth Weichel of this laboratory for her help during the preparation of this paper.

The Electro-Optic Effect in Uniaxial Crystals of the Type XH_2PO_4. II. Experimental

BRUCE H. BILLINGS
Research Laboratory, Baird Associates, Inc., Cambridge 38, Massachusetts

(Received June 22, 1949)

In this paper it is shown experimentally that when a voltage is applied to the Z direction to a basal section of ammonium dihydrogen phosphate the crystal becomes biaxial. The normal to the plate is the bisector of the axes and the plane of the axes is inclined at 45° to the crystallographic axes. The location of the axial plane is independent of the magnitude of the voltage. The retardation along the normal is directly proportional to the voltage and is independent of the thickness of the basal section. Photographs are shown of the interference pattern obtained with polarized light when voltage is applied to the crystal.

The behavior of the crystal plate with an alternating voltage is also described. Between polarizers the crystal can operate as a light valve both for short pulses as in a camera shutter or for complex sound waves as in a sound-on-film device. It is shown that with circular polarized light 74 percent sinusoidal modulation introduces only 3 percent harmonic distortion.

THE previous paper[1] discussed some theoretical aspects of the electro-optic effect in tetragonal scalenohedral crystals of the type XH_2PO_4. It was shown that the only large effects take place in a basal section with the field in the Z direction. For the field in the Z direction and for those members of the family in which a measurable retardation could be produced, the following conclusions were drawn:

(1) Uniaxial crystals of the form XH_2PO_4 become biaxial.
(2) The plane of the axes is independent of the magnitude of the voltage and is inclined at 45° to the crystallographic axis.
(3) At a given wave-length the retardation along the normal through the plate depends only on the voltage and is independent of the plate thickness.

The experimental techniques used in checking these conclusions will be discussed and the agreement between theory and experiment demonstrated in the case of the particular crystal ammonium dihydrogen phosphate[2] or ADP. This paper will also mention some of the problems connected with the use of an ADP basal section as an electro-optic shutter.

Perhaps the most important problem in connection with the use and measurements of the ADP electro-optic effect is the application of electrodes. Since the electric field must be applied parallel to the direction in which the light travels, the electrodes must be transparent. The electrode problem is made even more difficult by the physical characteristics of the crystal: Not only is the material brittle but it is water soluble and slightly hygroscopic.

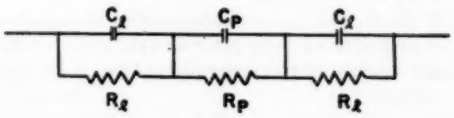

FIG. 1. Equivalent circuit of the laminated ADP crystal. C_l and C_p are the capacities of the laminating compound and the crystal. R_l and R_p are the resistance of the laminating compound and the crystal.

For the measurements which were made by Dr. Hans Jaffe beginning in 1942 at the Brush Development Company, glycerin was originally used as a conducting electrode. This had the disadvantages associated with the use of liquids and though the resulting electrodes were satisfactory electrically they were not by nature permanent.

At the Polaroid Corporation in 1943 evaporated gold was used as an electrode material. The gold layer which was used had a resistance of the order of 1000 ohms per square. Contact was made to this evaporated gold layer through an evaporated gold ring which was much heavier and had a resistance of only a few ohms per square. The disadvantage of this scheme is the loss of light at the gold surface. No matter how thin the evaporated layer, the reflection loss is still so great that more transparent electrodes would be desirable. An attempt to achieve greater transparency was made by Brian O'Brien, Jr. and Lt. Cmdr. G. A. Long at the University of Rochester in 1944. They used an evaporated grid of gold and relied on the surface conductivity of the PN to make the field uniform.

Most recent electrodes have been made using the Pittsburgh Plate Glass Nesa material. This material is a conducting layer of stannous oxide which is formed on glass. The material has no absorption in the visible part of the spectrum and for resistances of the order of 10,000 ohms per square the layers are so thin that the reflection losses are not much greater than the losses of the glass itself.

The optical quality of the Nesa is much better in thin coats than in very thick coats. One characteristic is haziness and non-uniformity that is apparent in thicknesses having resistances of a few 100 ohms/square. In the thicknesses giving 15,000 to 25,000 ohms/square this haziness disappears and the coat seems more uni-

[1] Bruce H. Billings, J. Opt. Soc. Am. 39, 797 (1949).
[2] Made by The Brush Development Company, Cleveland, Ohio, under the trademark "PN."

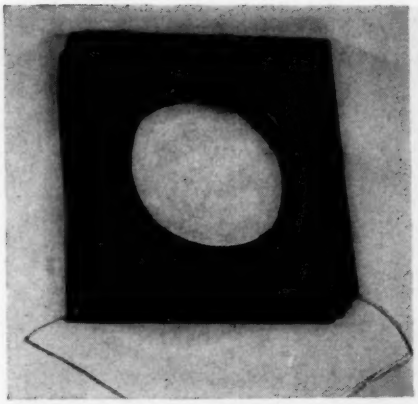

FIG. 2. ADP plate laminated between plates of Nesa-coated glass.

form. In both these and thick coats the glass is often severely strained. Selected pieces may be found which are strain-free. The surfaces are not flat and seem to depart from flatness by as many as 5–10 fringes in a 2-square.

The Nesa electrode introduces problems which are avoided with the use of glycerin or evaporated gold. In these other electrodes the crystal is in intimate contact with the conducting material. With the glass Nesa electrodes there is a thin film of non-conducting material between the crystal and the electrodes. If it were possible to get Nesa-coated plates which were optically flat, optical contact might be achieved between the Nesa and the crystal. Unfortunately, the method of preparing the conducting layer on the glass is of such a drastic nature that the glass is distorted. This makes optical contact out of the question.

Because of the extreme fragility of the PN crystal, an obvious technique is to fill the space between the crystal and the electrodes with a laminating compound such as Canada balsam or butyl-methacrylate. The glass electrode would then serve as a protective mount for the crystal, In this situation the voltage which actually appears across the crystal is a function of the characteristics of the laminating layer. The complete equivalent circuit of the ADP-Nesa glass sandwich is shown in Fig. 1. This circuit is applicable only at frequencies below about two megacycles. Below 10 kc the d.c. values of the constants can be used. Since in practice the Nesa plates have low resistance contacts along one edge only, the equivalent circuit at higher frequencies cannot be considered as made up of lumped elements. The new equivalent circuit will have the electrical properties of a transmission line. A detailed examination of the frequency response of the electro-optic effect in ADP and the significance of the electrodes at very high frequencies will be considered in a later paper.

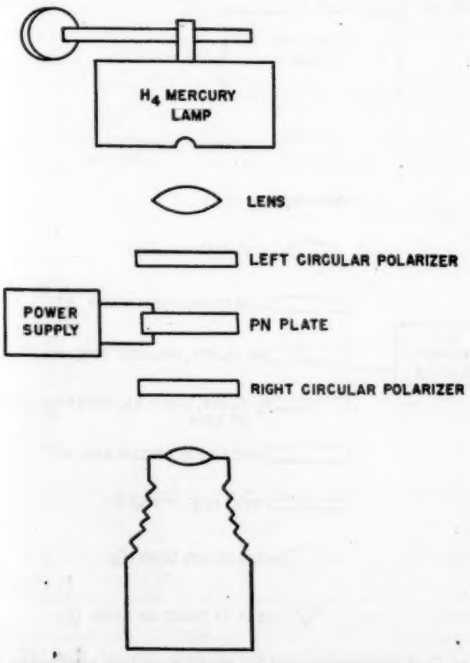

FIG. 3. Sketch of arrangement used to study the interference pattern of the ADP crystal with voltage applied.

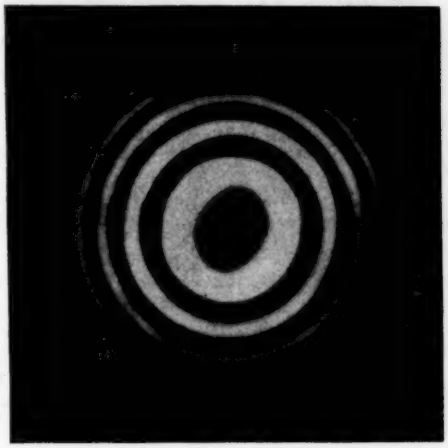

FIG. 4. Pattern observed at 546 mμ with basal section of ADP between crossed circular polarizers. Plate 0.31 cm thick.

At zero frequency the potential is divided according to the resistances of the ADP and laminating layers and is independent of the capacities. At a.c. frequencies the potential across the ADP can be written by inspection of the equivalent circuit and is:

$$V=\frac{V_0}{1+2(R_l/R_p)\left(\frac{1+j\omega R_pC_p}{1+j\omega R_lC_l}\right)}, \qquad (1)$$

FIG. 5. Pattern observed at 546 mμ with basal section of ADP 0.31 cm thick between crossed circular polarizers. 9600 volts on plate.

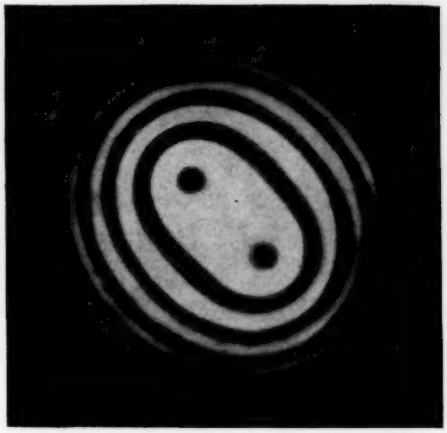

FIG. 6. Pattern observed at 546 mμ with basal section of ADP 0.31 cm thick between crossed circular polarizers. 4500 volts on plate.

for a condenser dielectric the quantity f_x, defined by

$$\frac{1}{2\pi R_x C_x}=\frac{1.80\times 10^{12}}{\rho_x K_x}=f_x, \tag{2}$$

where ρ and K are the volume resistivity (ohms-cm) and dielectric constant, respectively, is characteristic of the material and independent of the dimensions.

Equation (1) can be rewritten:

$$\left|\frac{V}{V_0}\right|^2=\frac{R_p^2(f_l^2+f^2)}{f_l^2(R_p+2R_l)^2+f^2\left(R_p+2R_l\frac{f_l}{f_p}\right)^2}. \tag{3}$$

According to this equation there are three different types of frequency response for the laminated ADP crystal.

Case I. $f_l=f_p$.

In this case the a.c. response is the same as the response to d.c. This case is unlikely to occur.

Case II. $f_l>f_p$.

In this case the response as a function of frequency drops for frequencies larger than:

$$f=f_l\frac{R_p+2R_l}{R_p+2R_l\frac{f_l}{f_p}}, \tag{4}$$

and then becomes constant for frequencies greater than $f=f_l$.

Case III. $f_l<f_p$.

This case is most likely to be realized in practice. Here the response climbs at $f>f_l$ and then becomes constant for frequencies greater than:

$$f=f_l\frac{R_p+2R_l}{R_p+2R_l\frac{f_l}{f_p}}. \tag{5}$$

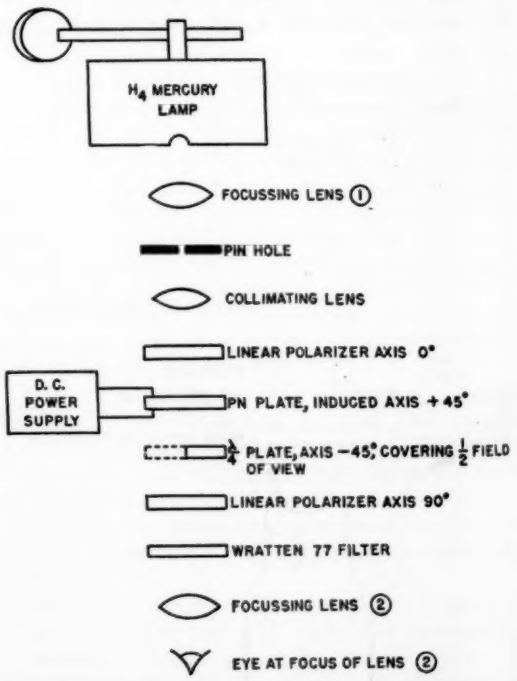
H4 MERCURY LAMP
FOCUSSING LENS ①
PIN HOLE
COLLIMATING LENS
LINEAR POLARIZER AXIS 0°
D. C. POWER SUPPLY
PN PLATE, INDUCED AXIS +45°
¼ PLATE, AXIS −45°, COVERING ½ FIELD OF VIEW
LINEAR POLARIZER AXIS 90°
WRATTEN 77 FILTER
FOCUSSING LENS ②
EYE AT FOCUS OF LENS ②

FIG. 7. Arrangement used for showing voltage required to produce the same normal retardation in basal sections of ADP of different thicknesses.

The values for ADP and for a typical laminating plastic such as plasticized polyvinyl butyral are:

PN: $f_p \sim 16$ c.p.s.
Plastic: $f_l \sim 3.2$ c.p.s.

For this particular laminating material the crystal will not respond adequately to frequencies below 3.2 cycles per second.

The d.c. response of such a lamination may be very poor. The crystal has a resistivity of 7800 megohm-cm and a typical plate will have a resistance of 300 megohms. A typical laminating layer will have a resistance of 500 megohms or higher. With such a combination only about a quarter of the applied voltage will appear across the crystal.

The improvement is so great for frequencies higher than the characteristic frequency of the lamination that for any a.c. work such a lamination is quite satisfactory.

One interesting consequence of the large crystal resistance is the slow rate of discharge if no low resistance shunt is present in the power supply. If d.c. is applied and the power supply disconnected from the crystal, the retardation remains unchanged. It is thus necessary in high speed pulse application to make provision for discharging the crystal at the end of a pulse as well as for charging it at the beginning.

Another set of problems arise because of the high voltage necessary to produce a half-wave retardation. For green light approximately 9500 volts is required to give one half-wave. The d.c. breakdown potential of a crystal 0.4 mm thick is above 25 kv/mm, but much lower values are found for a.c. potentials.[3]

For thicknesses of 0.080 in. which are used in the Baird Associates, Inc., cells, 9500 volts is well below the d.c. breakdown figure. It is necessary, however, to take special precautions to avoid breakdown around the edge of the crystal. Accordingly, the laminations are made so that the crystal projects well beyond the electrodes.

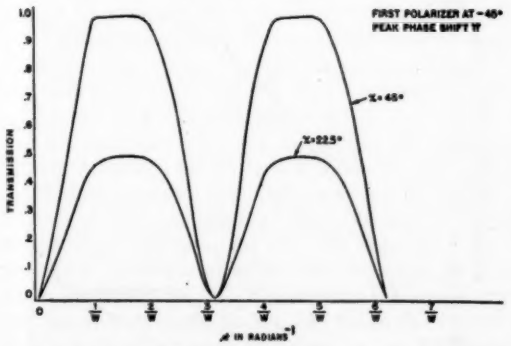

FIG. 8. ADP basal section transmission in collimated light as a function of time when a steady alternating current is applied. The crystal is between linear polarizers. One polarizer is at −45° to the plane of the induced axis, the other at +45° to the plane of the induced axis, the other at +45° and 22½° to the plane of the induced axes. Voltage peak gives ½-wave retardation.

[3] Personal communication from Dr. Hans Jaffe of The Brush Development Company, Cleveland, Ohio.

TABLE I. Voltage to produce retardation of approximately ½λ in crystals of different thicknesses.

Thickness	Voltage
0.08 in.	2000±200
0.25 in.	2000±200

TABLE II. Half-angular field to 1 percent and 0.1 percent transmission points for PN crystals of different thicknesses.

Thickness	Intensity	
	1.0%	0.1%
0.080 in.	2°20′	44′
0.250 in.	1°20′	26′

To check the conclusions of the analysis of the previous paper it is necessary to observe the interference pattern when voltage is applied to the crystal. These observations can be made either visually or with photographs. In general they can be made only with d.c. on the crystal and, from the data given above, a plastic lamination has too high a resistivity. Accordingly, the low resistance laminations were made with oleic acid which was suggested by Dr. Elkan R. Blout of Polaroid Corporation.

A typical ADP plate laminated between Nesa plates is shown in Fig. 2. In Fig. 3 is a sketch of the arrangement used in studying the behavior of the crystals as voltage is applied. The retardation through the plate is described by the equation:

$$\Gamma = -\frac{d}{\lambda}\frac{\omega^2-\epsilon^2}{\epsilon^2\omega}\sin^2 i, \qquad (6)$$

where λ is the wave-length, d the crystal thickness, ω and ϵ the ordinary and extraordinary indices, i is the angle of incidence, and Γ the retardation. When the

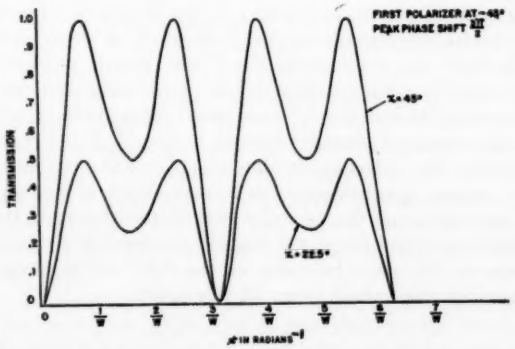

FIG. 9. ADP basal section transmission in collimated light as a function of time when a steady alternating current is applied. The crystal is between linear polarizers. One polarizer is at −45 to the plane of the induced axis, the other at +45° and 22½° to the plane of the induced axes. Voltage peak gives ½-waves retardation.

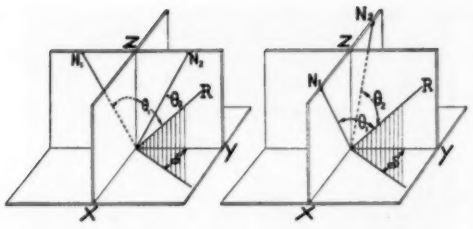

FIG. 10. Sketch showing that the angles between an oblique ray and optic axis are changed when the plane of the axis is rotated 90 degrees.

plate is placed between crossed circular polarizers it is opaque for integral values of Γ. The interference pattern thus consists of a series of rings and is shown in Fig. 4. It is typical of a basal section of any uniaxial crystal between crossed circular polarizers.

The first conclusion reached in the theoretical discussion in paper I was that the crystal would become biaxial with the axial plane at 45° to the crystallographic axis. The crystal section shown in Fig. 2 is approximately square and has edges which are parallel to the crystallographic axes. Accordingly, the biaxial interference pattern formed when voltage is applied should be at 45° to the edges of the crystal plate. This turns out to be the case. In Fig. 5 is shown the pattern obtained when 9600 volts is put across the crystal. In Fig. 6 is the pattern obtained with 4500 volts. The crystallographic axes were here kept parallel to the edge of the photograph. The optic axes remained in the same plane as the voltage was varied.

An experimental test was also made on the effect of crystal thickness on the voltage required for a given retardation.

The crystals were laminated with Nesa glass and oleic acid. The measurements were made with an arrangement devised by Mr. Sage of this laboratory. This is shown in Fig. 7. The crystal to be tested was oriented so that its induced axes would be at 45° to a linear polarizer. On the opposite side of the crystal was placed a strip of plastic of about $\frac{1}{4}$ λ-retardation at 546 mμ with its fast axes parallel to the slow axis of the crystal. Beyond the retardation sheet was placed a second polarizer at right angles to the first. Collimated light from an H4 mercury arc was passed through the system and examined visually through a lens. A Baird Associates, Inc., interference filter was placed in the system to restrict measurements to a wave-length of 546 mμ. The voltage on the crystal was then adjusted until the light passing through the retardation sheet and crystal was of the same intensity as the light which passed through the crystal alone. At this point

$$\Gamma_c=\frac{\Gamma_s}{2}, \tag{7}$$

when Γ_c is the retardation through the crystal and the retardation of the retardation sheet. With this arrangement it was not necessary to know the actual retardation of the crystal. Balance in the light beams is achieved at one value of the retardation and for this to be independent of thickness it must take place at the same voltage for crystals of different thickness. The results of the measurements are shown in Table I and it is apparent that the predicted independence of retardation and thickness is realized experimentally.

ANGULAR FIELD OF VIEW

When the crystal is used in an instrument as a light valve, there are several other characteristics that are involved in its behavior. One of the most important is the angular field of view. Since the crystal with zero applied voltage is uniaxial, the retardation is zero only for light parallel to the optic axis. For light inclined to the axis the retardation is finite and linearly polarized light is rendered sufficiently elliptical to penetrate the second polarizer. Superficially it might appear that the angular field of view under such conditions should be zero. Actually the angular field is finite as a practical limit can be set on the amount of light to be passed when the shutter is closed.

The equation which describes the light transmitted through a basal section of a uniaxial crystal between crossed circular polarizers can be written:

$$I=\tfrac{1}{2}\sin^2\left(\frac{d}{\lambda}\frac{\omega^2-\epsilon^2}{2\omega\epsilon^2}\sin^2 i\right). \tag{8}$$

If the light leak at the edge of the field is set at, for

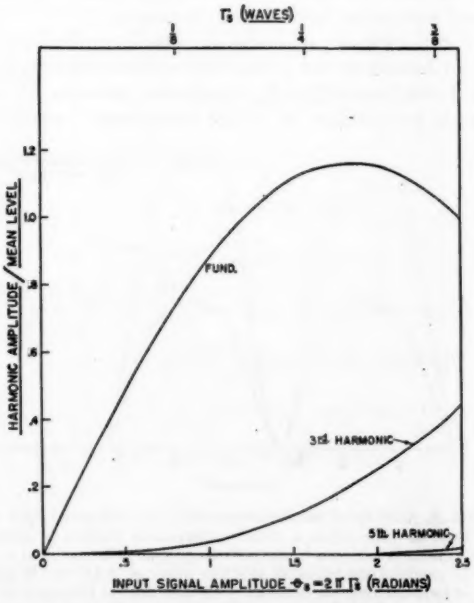

FIG. 11. Relative value of fundamental third harmonic and fifth harmonic in transmitted light when crystal is modulated at various amplitudes with a sine wave.

instance, 1 percent of the center transmission, this equation gives a value for the field of view as a function of the crystal thickness. It shows that the field of view is inversely proportional to the square root of the crystal thickness. In Table II is shown values of the field for different thickness systems with an allowance of 0.1 percent and of 1 percent light leak.

It might appear that any light leak would be intolerable especially in most photographic applications. This is not true if a simple mechanical shutter is put in series with the crystal shutter. The mechanical shutter is opened just as the crystal shutter is to be used. The tolerable light leak will then depend on the brightness of the scene and the ratio of the time of opening of the crystal shutter to the time of opening of the mechanical shutter.

The field of view of the open shutter is less critical than the field of the closed shutter. Since the whole area which is usefully closed becomes transparent when the crystal is open, no investigation needs to be made here of this case.

OPERATION WITH ALTERNATING CURRENT

We shall consider first the case in which the crystal is placed between crossed linear polarizers. The modulation voltage will be assumed to be sinusoidal. With appropriate frequency response curves it will then be possible to show the behavior of the crystal for any modulation function. The crystal can be considered as a simple retardation plate between crossed polarizers. The intensity of the light through the combination can be written:

$$I=\tfrac{1}{2}[1-\cos(\delta_0 \sin\omega t)]\sin^2(2\chi), \tag{9}$$

where χ is the angle between induced crystal axes and the first polarizer, ω is the angular frequency of the a.c. voltage, and δ_0 is the retardation of the crystal at peak voltage. A plot of this equation as a function of the time t is shown in Fig. 8. This is plotted with a value of for δ_0. If δ_0 is more than π, the peaks are split. A plot of this splitting is shown in Fig. 9. This figure is drawn for the case then the peak phase shift is $3\pi/2$.

If the light goes through the crystal at other than normal incidence, the pattern is more complex. When the field changes sign the optic axes open in a plane at 90° to the original direction. The retardation depends on the angles between the ray direction and the two optic axes. These angles will be different when the plane of the axes is shifted 90°. The relation between the ray and the axes is shown in Fig. 10. When the crystal is placed between crossed polarizers in an alternating field, the intensity will thus be different for the two parts of the cycle.

APPLICATIONS

From these various experiments both with direct current and with alternating current it is possible to see many of the possible applications of the electro-optic crystal. The crystals can be used as a light valve for putting sound on film. This has been discussed by Gotschall.[4]

The device has several advantages over the conventional sound on film light valve. It is easily adaptable to portable operations and is essentially non-microphonic.

One problem in connection with this application is the question of the linearity of the crystal. The change in intensity of the light which passes through the PN must follow exactly the change in voltage on the crystal. The analysis below was carried out by Mr. R. O'B. Carpenter of this laboratory. The intensity along the normal to the crystal between parallel polarizers can be written:

$$I/I_0=\tfrac{1}{2}\cos^2\pi\Gamma, \tag{10}$$

where Γ is the retardation of the crystal. This retardation is induced by the voltage. In addition to this electro-optic retardation, a retardation can be produced by putting in series with the crystal a fixed retardation sheet.

Consider that such a fixed plate of retardation Γ_0 is in series with the crystal and that a sinusoidal voltage is applied to the PN. In this case

$$I/I_0=\tfrac{1}{4}[1+\cos2\pi(\Gamma_0+\Gamma_s \sin\omega t)], \tag{11}$$

where Γ_s is the peak retardation produced in the modulation elements by the applied voltage and ω is the angular frequency of this voltage. Equation (11) can be expanded to

$$I/I_0=\tfrac{1}{4}[1+\cos\theta_0\cos(\theta_s \sin\omega t)-\sin\theta_0\sin(\theta_s \sin\omega t)], \tag{12}$$

where $\theta=2\pi\Gamma$.

This can be rewritten in terms of Bessel functions as

$$I/I_0=\tfrac{1}{4}\Big[1+\cos\theta_0\Big\{J_0(\theta_s)+J_2(\theta_s)\cos2\omega t + J_4(\theta_s)\cos4\omega t+\cdots\Big\} -\sin\theta_0\Big\{J_1(\theta_s)\sin\omega t+J_3(\theta_s)\sin3\omega t+\cdots\Big\}\Big]. \tag{13}$$

This gives the amplitudes of the various harmonics as a function of the operating point and signal amplitude.

At $\theta_0=\pi/2$($\Gamma_0=\frac{1}{4}$ wave) all even harmonics disappear. At $\theta_0=0$ all odd harmonics including the fundamentals disappear. In Fig. 11 is plotted the amplitude of the various harmonics relative to maximum transmission for the proper operating point $\theta_0=\pi/2$. With a maximum signal of $\frac{1}{8}$ wave the percentage modulation is 74 percent and the third harmonic is less than 3 percent of the fundamental—all other harmonics being negligible.

The crystal also can be used as a high speed camera shutter.[5] This application is restricted by the narrow angular field of view. For very short exposure times film speeds are such that only the illumination of

[4] G. D. Gotschall, J. Soc. Mot. Pict. Eng. **51**, 13 (1948).
[5] Roy Gunter, J. Opt. Soc. Am. **38**, 419 (1948).

special flash discharge tubes is sufficiently intense. A high speed shutter might appear unnecessary. Professor Edgerton has pointed out, however, that the light pulse from these lamps has a tail which is troublesome for certain problems. A crystal shutter could effectively remove this tail.

Besides being used as a high speed shutter the crystal appears useful in the lower speed range because of the precision with which the exposure can be set.

In the tunable birefringent filter[6] this crystal is one of the most promising tuning elements and makes possible extremely rapid shifts of the filter passband.

In conclusion, the author wishes to express appreciation to Mr. Stanley J. Sage, Mr. Edgar E. Barr, and Mr. George Economou of this laboratory for their help during the preparation of this paper.

[6] B. H. Billings, J. Opt. Soc. Am. 37, 738–746 (1947).

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 39, NUMBER 10 OCTOBER, 1949

Visual Sensitivities to Combined Chromaticity and Luminance Differences*

W. R. J. BROWN** AND D. L. MACADAM
Kodak Park Works, Eastman Kodak Company, Rochester 4, New York

(Received June 16, 1949)

Color discrimination for the general case of combined chromaticity and luminance differences has been investigated by analysis of the errors of visual trichromatic colorimetry. Formulas based on those devised by Silberstein have been used to determine the coefficients and axes of ellipsoids in color space from the data. Methods for transforming the results to other coordinate systems are presented and used to give the results in terms of both the ICI tristimulus X, Y, Z values and the ICI x, y, chromaticity values. Most of the results are nearly symmetrical above and below the chromaticity (constant luminance) cross section of each ellipsoid. Those cross sections are comparable with results previously published, and the agreement is satisfactory. The influence on color discrimination of changing the level of luminance has been studied for three colors. The ratio between the standard of deviation of each primary and its amount seems to be a function of only the luminance contributed by that primary to the color mixture.

I. INTRODUCTION

THE purpose of this investigation is to determine standard deviations of color-matchings for representative colors scattered throughout the color domain. Standard deviation of color-matching is directly related to the corresponding just-noticeable difference of colors. The standard ICI primaries allow all colors to be specified in terms of three numbers. The selection of the ICI primary colors was based on considerations of computational convenience. Very little consideration was given to spacing points representing different colors in accordance with the perceptibility of their differences throughout the color domain. One reason why this was not taken into consideration was that there was only fragmentary knowledge of the relations between specifications and noticeability of the corresponding color differences.

In terms of the ICI primaries, it might be asked, "If the just-perceptible difference in appearance of greens corresponds to a change of coordinates of two units, then will two units also indicate a just-perceptible difference of blue colors?" The answer is that they will not. It is the purpose of this investigation to determine the perceptual magnitudes associated with variations of ICI coordinates throughout the domain of colors.

The ICI specifications of colors are in extensive use. In order adequately to specify a tolerance in any color in terms of ICI coordinates, the just-noticeable differences of dominant wave-length purity, and luminance must be known. These differential thresholds vary from color to color over a wide range of values. Many investigations have been undertaken, the purpose of which was to determine the variation of just-noticeable differences. The most complete surveys were carried out by Wright and MacAdam.[1] In these researches, the sensitivity of the eye to changes of color was measured at constant luminance. MacAdam plotted the results on the chromaticity diagram in terms of the standard deviation of color-matching in several directions about selected colors. The resulting figures formed closed curves on the diagram. Silberstein and MacAdam[2] observed that the distribution of errors in the color-matching in the above experiment was statistically normal. From this they deduced that the curves should have the form of ellipses, as it appeared they had. In the same paper they suggested that if the matches were "free," i.e., not confined to chromaticity variations, then the surfaces in color space representing the standard deviations of color-matching would be ellipsoids.

Silberstein[3] assumed that the probability of making a

* Communication No. 1268 of the Kodak Research Laboratories.
** Submitted to Physics Department of University of Rochester in partial fulfillment of requirements for degree of Master of Science.

[1] W. D. Wright, Proc. Phys. Soc. 53, 93–112 (1941); D. L. MacAdam, J. Opt. Soc. Am. 32, 247–274 (1942).
[2] L. Silberstein and D. L. MacAdam, J. Opt. Soc. Am. 35, 32–39 (1945).
[3] L. Silberstein, Phil. Mag., Ser. 7, 37, 126–144 (1946).

match that falls within a specific region of color space, near the standard aimed at, is not changed by any change of primaries. From this he proved the standard deviation figures to be ellipsoids and gave formulas from which the coefficients of the ellipsoids could be determined. These ellipsoids can be imagined as formed by plotting the standard deviation of color-matching as a radius of the surface. The greater the standard deviation in a particular direction, the greater will be the radius in that direction. The coefficients given by Silberstein's formulas determine directly the orientation and size of the ellipsoids in color space, eliminating the necessity of determining a large number of separate radii. The purpose of this investigation will be realized by evaluating these coefficients for a large number of representative color centers. Once determined, these ellipsoids completely specify the differential sensitivity of the eye throughout the domain of colors.

II. THEORY

In order to calculate the size and orientation of ellipsoids formed by the standard deviation of color-matching, it may be assumed that the distribution of matchings in any direction is normal. This assumption has been verified experimentally by MacAdam's research, previously mentioned, and in the present investigation.

The probability of a color-match falling at a position, $x_1x_2x_3$, relative to a central color, XYZ, is given by

$$P=A\exp-(c_{11}x_1^2+c_{12}x_1x_2+c_{13}x_1x_3+c_{21}x_2x_1+c_{22}x_2^2+c_{23}x_2x_3+c_{31}x_3x_1+c_{32}x_3x_2+c_{33}x_3^2)\cdot dx_1dx_2dx_3,$$

or more compactly,***

$$P=A\exp(-c_{ik}x_ix_k)\cdot dx_1dx_2dx_3.$$

This is the well-known normal law for three variables. The normal frequency function may be written

$$f(x_1, x_2, x_3)=A\exp(-c_{ik}x_ix_k).$$

The constant A may be evaluated by considering that the probability that a single color-match falls somewhere within color space is unity:

$$1=A\int\int\int_{-\infty}^{+\infty}\exp(-c_{ik}x_ix_k)\cdot dx_1dx_2dx_3.$$

By evaluating this integral, it can be shown that

$$A=(D/\pi^3)^{\frac{1}{2}},$$

where

$$D=\begin{vmatrix}c_{11} & c_{12} & c_{13}\\ c_{21} & c_{22} & c_{23}\\ c_{31} & c_{32} & c_{33}\end{vmatrix}=|c_{ik}|$$

is the determinant of the coefficients of the exponent.

*** In this expression, and throughout this article, whenever an index letter (such as both i and k in this case) appears twice in a product term, it is to be understood that all possible products that can be formed by giving each such letter all allowable numerical values (1, 2, 3, in this article) are to be added together.

Upon substitution, the frequency function becomes

$$f(x_1, x_2, x_3)=(|c_{ik}|/\pi^3)^{\frac{1}{2}}\exp(-c_{ik}x_ix_k).$$

Consider a number (n) of free trichromatic color-matchings $(x_1'x_2'x_3')$, $(x_1''x_2''x_3'')$, $\cdots$, $(x_1^nx_2^nx_3^n)$. Now the most probable values of the nine coefficients of the exponent, c_{11}, c_{12}, c_{13}, c_{21}, c_{22}, c_{23}, c_{31}, c_{32}, c_{33} are those that will yield a maximum value of the product of $f_1\cdot f_2\cdot f_3\cdots f_n$, where $f_j=f(x_1^jx_2^jx_3^j)$. This is equivalent to requiring the sum,

$$\sum_{j=1}^{n}\log f_j,$$

to have the maximum value permitted by variation of c_{ik}. In order that this be a maximum, all the derivatives of $\sum\log f_j$ with respect to c_{11}, c_{12}, c_{13}, $\cdots$, c_{33} must be zero. Differentiating, we obtain

$$(1/n)\partial(\sum\log f_j)/\partial c_{ik}=\partial(\tfrac{1}{2}\log|c_{ik}|-c_{ik}\sum x_ix_k/n)/\partial c_{ik}=0$$
$$=\mathbf{C}_{ik}/2|c_{ik}|-\sum x_ix_k/n=0.$$

In the last equation, above, the boldface capital, $\mathbf{C}_{ik}$, denotes the co-factor of c_{ik} in the determinant, $|c_{ik}|$.

It is to be noted that since $x_1x_2x_3$ represent the difference of a match from a central color,

$$(\sum x_1^2/n)^{\frac{1}{2}},\quad(\sum x_2^2/n)^{\frac{1}{2}},\quad\text{and}\quad(\sum x_3^2/n)^{\frac{1}{2}}$$

represent the standard deviations of the color-matching for each primary. Also,

$$\sum x_ix_k\equiv\sum x_kx_i,$$

because an ordinary multiplication is a commutative process. This makes three pairs of components of the tensor, c_{ik}, equal. The frequency function becomes

$$f(x_1, x_2, x_3)=(|c_{ik}|/\pi^3)^{\frac{1}{2}}\exp(-(c_{11}x_1^2+2c_{12}x_1x_2+c_{22}x_2^2+2c_{23}x_2x_3+c_{33}x_3^2+2c_{13}x_1x_3)).$$

The term $\mathbf{C}_{ik}/|c_{ik}|$ in the equations derived above, is an element of the conjugate tensor, c^{ik}; therefore, $c^{ik}=2\sum x_ix_k/n$. The determination of the desired c_{ik}, therefore, reduces to a determination of the tensor conjugate to $2\sum x_ix_k/n$, that is, $c_{ik}=\mathbf{C}^{ik}/|c^{ik}|$ where, again the boldface capital, $\mathbf{C}^{ik}$, denotes the co-factor of c^{ik} in the determinant, $|c^{ik}|$.

The terms $\sum x_ix_k/n$ may be determined from a group of trichromatic color-matchings. These may be doubled to find the terms of the tensor, c^{ik}, from which the six desired values of the inverse tensor, c_{ik}, can be found. In order to determine $\sum x_ix_k/n$ from the experimental data, one other step is involved. The intensity in the colorimeter is varied by the inverse square law, hence $X_i=K/L_i^2$, where L_i is the distance of the lamp from the surface it illuminates, and

$$x_i=\Delta X_i=-2X_i\Delta L_i/L_i=-2X_i\Delta S_i/L_i,$$

where scale reading S_i may be from some arbitrary

origin, not necessarily zero when L_i is zero. Therefore,

$$\sum x_i x_k/n=(4X_iX_k/L_iL_k)(\sum\Delta S_i\Delta S_k/n)$$

and

$$c^{ik}=(8X_iX_k/L_iL_k)(\sum\Delta S_i\Delta S_k/n),$$

where

$X_1=R$ is the amount of the red primary in the mixture;
$X_2=G$ is the amount of the green primary in the mixture;
$X_3=B$ is the amount of the blue primary in the mixture;
n is the number of observations; and ΔS_i is the variation of each setting from the mean value.

Numbers having inconvenient magnitudes, and rejection errors, can be avoided in the inversion of the tensor, c^{ik}, by factoring out the term $(c^{11})^{\frac{1}{2}}$ from the first row and from the first column, $(c^{22})^{\frac{1}{2}}$ from the second row and second column, and $(c^{33})^{\frac{1}{2}}$ from the third row and third column. This yields a reduced tensor having unit values on the main diagonals, and quantities of absolute magnitude less than unity off the main diagonal. Those quantities, which may be either positive or negative, are correlation coefficients, ρ_{ik}, measuring the degree to which a variation of one primary makes an error of another primary more or less conspicuous. The factors removed from the rows and columns are each $\sqrt{2}$ times the standard deviation, σ_r, σ_g, σ_b, of the corresponding primary, i.e.,

$$(c^{11})^{\frac{1}{2}}=\sqrt{2}\sigma_r,\quad (c^{22})^{\frac{1}{2}}=\sqrt{2}\sigma_g,\quad \text{and}\quad (c^{33})^{\frac{1}{2}}=\sqrt{2}\sigma_b.$$

The reduced tensor may be denoted

$$\rho_{ik}=\begin{matrix}1 & \rho_{rg} & \rho_{rb}\\ \rho_{rg} & 1 & \rho_{gb}\\ \rho_{rb} & \rho_{gb} & 1.\end{matrix}$$

The correlation coefficients can be computed directly from the deviations of the scale readings from their means. Thus,

$$\rho_{rg}=\sum_1^n \Delta S_r\cdot\Delta S_g/(\sum\Delta S_r^2)^{\frac{1}{2}}\cdot(\sum\Delta S_g^2)^{\frac{1}{2}}.$$

Similarly, the standard deviations of the primaries can be computed from the standard deviations of the scale readings. Thus,

$$\sigma_r=2[\bar{R}/(\bar{S}_r+455)]\cdot[\sum\Delta S_r^2/n]^{\frac{1}{2}},$$

where $\bar{S}_r$ is the average value of the scale readings of the R primary, expressed in millimeters; $\bar{R}$ is the corresponding intensity of the R primary: $\bar{R}=10^6T_r/(\bar{S}_r+455)^2$; T_r is the transmittance of the neutral filter, if any, used in the red beam; S_r is each deviation (in millimeters) of the scale reading from the average; and n is the number of readings.

The inverse ρ^{ik} of the reduced tensor is easily found by use of the formula, $\rho^{ik}=P_{ik}/|\rho_{ik}|$, where P_{ik} is the co-factor, or determinant of the minor, of ρ_{ik} in the determinant $|\rho_{ik}|$. The desired inversed tensor is then given by

$$c_{ik}=\rho^{ik}/2\sigma_i\sigma_k.$$

The c_{ik}'s computed from the observational data as just indicated specify visual sensitivity to color difference expressed in terms of the instrument primaries, R, G, and B. Those primaries are not of general interest or importance, and in fact they were changed twice during the investigation. It is much more important to specify visual sensitivities to color differences expressed in terms of the standard ICI primaries, X, Y and Z, or in terms of the standard chromaticity coordinates x, y, and luminance, Y. A color difference may be expressed equally well in any of the three following forms:

ΔR, ΔG, ΔB of the instrument primaries
ΔX, ΔY, ΔZ of the standard ICI primaries
Δx, Δy, ΔY of the chromaticity and luminance.

The noticeability of the color difference may be measured, in multiples of the standard deviation of color-matching, by S, which can be computed by any one of the following formulas:

$$\begin{aligned}\Delta S^2&=c_{11}\Delta R^2+2c_{12}\Delta R\Delta G+c_{22}\Delta G^2+2c_{23}\Delta G\Delta B\\&\qquad+c_{33}\Delta B^2+2c_{13}\Delta B\Delta R\\&=G_{11}\Delta X^2+2G_{12}\Delta X\Delta Y+G_{22}\Delta Y^2+2G_{23}\Delta Y\Delta Z\\&\qquad+G_{33}\Delta Z^2+2G_{13}\Delta X\Delta Z\\&=g_{11}'\Delta x^2+2g_{12}'\Delta x\Delta y+g_{22}'\Delta y^2+2g_{23}'\Delta y\Delta Y\\&\qquad+g_{33}'\Delta Y^2+2g_{13}'\Delta x\Delta Y.\end{aligned}$$

The quantity c_{ik} is a second-rank covariant tensor. In order to transform a covariant tensor to another coordinate system, such as X, Y, Z, the following relation is used:

$$G_{jl}=(\partial x_i/\partial x_j')(\partial x_k/\partial x_l')\cdot c_{ik},$$

where

$$\begin{aligned}&x_i,\ x_k=R,\ G,\ B\\&x_j',\ x_l'=X,\ Y,\ Z.\end{aligned}$$

The values of $\partial x_i/\partial x_j'$ and $\partial x_k/\partial x_l'$ are obtained by differentiating the transformation equations, which in the present case are

$$x_i=a_{ij}x_j',$$

or written out:

$$\begin{aligned}R&=a_{11}X+a_{12}Y+a_{13}Z\\G&=a_{21}X+a_{22}Y+a_{23}Z\\B&=a_{31}X+a_{32}Y+a_{33}Z.\end{aligned}$$

For example,

$$\partial x_1/\partial x_1'=a_{11},\quad \partial x_2/\partial x_1'=a_{21},\quad \partial x_3/\partial x_1'=a_{31}.$$

Therefore,

$$\begin{aligned}(\partial x_1/\partial x_1')(\partial x_2/\partial x_1')&=a_{11}a_{21}\\(\partial x_2/\partial x_1')(\partial x_3/\partial x_1')&=a_{21}a_{31}\end{aligned}$$

and

$$(\partial x_1/\partial x_1')(\partial x_3/\partial x_1')=a_{11}a_{31}.$$

Writing out one equation in full, we obtained

$$\begin{aligned}G_{11}=&(\partial x_1/\partial x_1')(\partial x_1/\partial x_1')c_{11}+(\partial x_1/\partial x_1')(\partial x_2/\partial x_1')c_{12}\\&+(\partial x_1/\partial x_1')(\partial x_3/\partial x_1')c_{13}+(\partial x_2/\partial x_1')(\partial x_1/\partial x_1')c_{21}\\&+(\partial x_2/\partial x_1')(\partial x_2/\partial x_1')c_{22}+(\partial x_2/\partial x_1')(\partial x_3/\partial x_1')c_{23}\\&+(\partial x_3/\partial x_1')(\partial x_1/\partial x_1')c_{31}+(\partial x_3/\partial x_1')(\partial x_2/\partial x_1')c_{32}\\&+(\partial x_3/\partial x_1')(\partial x_3/\partial x_1')c_{33},\end{aligned}$$

or, noting that $c_{21}=c_{12}$, $c_{32}=c_{23}$, and $c_{31}=c_{13}$

$$G_{11}=a_{11}^2c_{11}+2a_{11}a_{21}c_{12}+a_{21}^2c_{22}+2a_{21}a_{31}c_{23}+a_{31}^2c_{33}+2a_{11}a_{31}c_{13}.$$

Proceeding similarly, we find that

$$G_{12}=a_{11}a_{12}c_{11}+(a_{11}a_{22}+a_{21}a_{12})c_{12}+a_{21}a_{22}c_{22}+(a_{21}a_{32}+a_{31}a_{22})c_{23}+a_{31}a_{32}c_{33}+(a_{11}a_{32}+a_{31}a_{12})c_{13}.$$

Substituting the values for the a_{ik} in the equations gives the tensor transformations directly. The coefficients of the ellipsoid in the standard color space may be found by substituting in these equations.

The presentation of the ellipsoids in terms of the ICI chromaticity coordinates, x, y, and Y is also useful. To find the transformation equations from X, Y, and Z to x, y, Y, the same transformation rule can be used as before, i.e.,

$$g_{jl'}=(\partial x_i/\partial x_j')(\partial x_k/\partial x_l')G_{ik}.$$

In this case, however, the transformation between X, Y, Z and x, y, Y is given by

$$\begin{aligned} x_1&=x_1'x_3'/x_2', && \text{i.e.,}\quad X=xY/y\\ x_2&=x_3', && \text{i.e.,}\quad Y=Y\\ x_3&=(1-x_1'-x_2')x_3'/x_2', && \text{i.e.,}\quad Z=(1-x-y)Y/y. \end{aligned}$$

By carrying out the indicated operations, we obtain

$$\begin{aligned} g_{11}'&=S^2(G_{11}+G_{33}-2G_{13})\\ g_{12}'&=S^2(-AG_{11}+BG_{33}+\{A-B\}G_{13})\\ g_{22}'&=S^2(A^2G_{11}+B^2G_{33}+2ABG_{13})\\ g_{23}'&=S(-A^2G_{11}-AG_{12}-BG_{23}-BCG_{33}-A\{B+C\}G_{13})\\ g_{33}'&=A^2G_{11}+2AG_{12}+G_{22}+2CG_{23}+C^2G_{33}+2ACG_{13}\\ g_{13}'&=S(AG_{11}+G_{12}-G_{23}-CG_{33}+\{C-A\}G_{13}) \end{aligned}$$

where

$$A=x/y,\quad B=(1-x)/y,\quad C=(1-x-y)/y$$

and

$$S=X+Y+Z=Y/y.$$

These equations determine the coefficients of the ellipsoids in terms of the coordinates, x, y, and Y.

In order to make the dimensions of ellipsoids in the Y direction of the same order of magnitude as their dimensions in the x and y directions, a scaling factor was found desirable. This scaling factor was arbitrarily inserted by dividing the terms in ΔY by kY, where k is a constant. In order to maintain the same measure of noticeability of color difference, ΔS, it is necessary to multiply the coefficients, g_{23}' and g_{13}' by the same factor, and g_{33}' by its square, i.e.,

$$\Delta S^2=g_{11}'\Delta x^2+2g_{12}'\Delta x\Delta y+g_{22}'\Delta y^2+(2g_{23}'kY)\Delta y(\Delta Y/kY)+(g_{33}'k^2Y^2)(\Delta Y/kY)^2+(2g_{13}'kY)\Delta x(\Delta Y/kY).$$

It may be noted that if $k=1$, this scaling factor changes the coordinates from x, y, Y to x, y, $\ln Y$, since $\Delta Y/Y=\Delta \ln Y$. Ellipsoids of more informative shape are obtained if k is about ten. If the value $k=11.53=5\times 2.306$ is chosen, so that $\ln Y/k=(\log_{10}Y)/5$, excessive elongation of the ellipsoids in the Y direction is avoided and

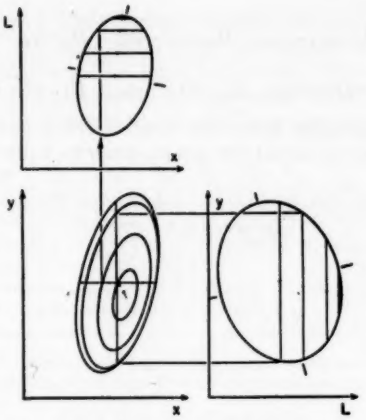

FIG. 1. Standard deviation ellipsoid, representing combined chromaticity and luminance discrimination, with constant-luminance cross sections showing displacement of center and of point of maximum luminance.

the coordinates are in a convenient form. Using the new coordinate, $l=0.2\log Y$, the equation of the ellipsoid becomes:

$$1=g_{11}'\Delta x^2+2g_{12}'\Delta x\Delta y+g_{22}'\Delta y^2+2kYg_{23}'\Delta y\Delta l+(kY)^2g_{33}'\Delta l^2+2kYg_{13}'\Delta x\Delta l.$$

The most convenient place to insert the scaling factor is in the transformation equations:

$$g_{11}=g_{11}',\quad g_{12}=g_{12}',\quad g_{22}=g_{22}',\quad g_{23}=kYg_{23}',\quad g_{33}=(kY)^2g_{33}',\quad g_{13}=kYg_{13}'.$$

These equations yield the coefficients of the ellipsoids after scaling. The equation of the ellipsoid then becomes

$$1=g_{11}\Delta x^2+2g_{12}\Delta x\Delta y+g_{22}\Delta y^2+2g_{23}\Delta y\Delta l+g_{33}\Delta l^2+2g_{13}\Delta x\Delta l.$$

It is possible to convert the discrimination tensor, c_{ik}, referring to the instrument primaries, directly into the tensor, g_{ik}, referring to the ICI chromaticity coordinates and luminance, Y (or, as we desire, $l=0.2\log Y$). In many cases, this direct transformation avoids serious rejection errors and the compounding of errors that are encountered in the two successive transformations previously indicated.

The direct transformation is given by

$$\begin{aligned} g_{11}&=S^2[D_1^2c_{11}+2D_1D_2c_{12}+D_2^2c_{22}+2D_2D_3c_{23}+D_3^2c_{33}+2D_1D_3c_{13}]\\ g_{12}&=-(S^2/y)[D_1E_1c_{11}+(D_1E_2+D_2E_1)c_{12}+D_2E_2c_{22}+(D_2E_3+D_3E_2)c_{23}+D_3E_3c_{33}+(D_1E_3+D_3E_1)c_{13}]\\ g_{22}&=(S^2/y^2)[E_1^2c_{11}+2E_1E_2c_{12}+E_2^2c_{22}+2E_2E_3c_{23}+E_3^2c_{33}+2E_1E_3c_{13}]\\ g_{23}&=-(kS/y)[E_1Rc_{11}+(E_1G+E_2R)c_{12}+E_2Gc_{22}+(E_2B+E_3G)c_{23}+E_3Bc_{33}+(E_1B+E_3R)c_{13}]\\ g_{33}&=k^2[R^2c_{11}+2RGc_{12}+G^2c_{22}+2GBc_{23}+B^2c_{33}+2RBc_{13}]\\ g_{13}&=kS[D_1Rc_{11}+(D_1G+D_2R)c_{12}+D_2Gc_{22}+(D_2B+D_3G)c_{23}+D_3Bc_{33}+(D_3R+D_1B)c_{13}], \end{aligned}$$

where

$$D_1=a_{11}-a_{13},\quad D_2=a_{21}-a_{23},\quad D_3=a_{31}-a_{33}$$

and

$$E_1=D_1x+a_{13},\quad E_2=D_2x+a_{23},\quad E_3=D_3x+a_{33}.$$

These formulas have been used to check and in some instances to correct the results given in Table V.

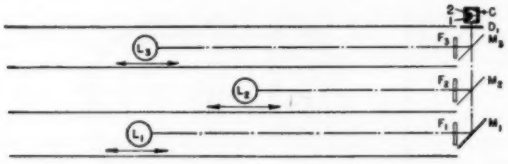

FIG. 2. Front view of colorimeter used to determine visual sensitivities for combined chromaticity and luminance differences.

When Δl (or ΔY itself) is zero, the equation of the discrimination ellipsoid reduces to the equation of the elliptical cross section in the chromaticity diagram for constant luminance. The values of g_{11}, g_{12}, g_{22} derived from the present investigation should therefore be directly comparable with those previously recorded for observations at constant luminance.[4]

When $\Delta x=\Delta y=0$, the equation reduces to

$$1=g_{33}\Delta l^2=g_{33}(\Delta Y/11.53Y)^2.$$

Since, in this case, ΔY denotes the standard deviation of luminance-matching in the absence of chromaticity difference, and since three times the standard deviation is approximately equal to a just-noticeable difference, $3\sigma Y/Y$ should be approximately equal to Weber's fraction, W. According to Hecht, Peskin, and Patt[5] this is nearly independent of chromaticity, and therefore g_{33} should be nearly independent of x and y. Since Weber's fraction is nearly independent of luminance, Y, at the levels of luminance (0.5 to 10 footlamberts) employed in this investigation, g_{33} should be nearly the same in all cases. Specifically,

$$g_{33}=1/(\Delta Y/11.53Y)^2\cong(3\times11.53)^2/W^2\cong1200/W^2,$$

when

$$\Delta x=\Delta y=0.$$

In order to construct models that represent the results in an informative way, it is necessary to determine the directions and lengths of the principal axis of the ellipsoids in some kind of colorimetric coordinate space. There would be little utility in doing this in the space defined by the instrument primaries, R, G, and B, since they are peculiar to this investigation and were changed twice during it. There would be more reason for representing them in the space defined directly by the standard ICI primaries, X, Y, Z. However, it appears much more useful and instructive to show the ellipsoids in the space defined by the rectangular ICI chromaticity diagram, x, y, and the logarithm of luminance ($\log Y$) plotted to one-fifth scale, perpendicular to that diagram. Colorimetric data are usually given in the form of chromaticity and some quantity related to luminance, Y. Use of ($\log Y$) eliminates all dependence of the results on arbitrary choice of units of luminance, and renders the results equally applicable to specifications in terms of luminous reflectance or any other photometric quantity.

That the cross sections in planes representing constant Y are comparable with the Nutting ellipses[6] has been pointed out, as has also the close relationship between the radii perpendicular to those planes and the Weber fraction. Whether or not luminance differences are masked or accentuated by some kinds of chromaticity variation is indicated by the inclination of the principal axis with respect to the planes of constant Y. In all of these respects, the x, y, $\log Y$ representation of the ellipsoids is more convenient than the X, Y, Z representation. In the latter representation, the ellipsoids would expand almost proportionally to luminance when various luminances of constant chromaticity are studied.[7] This dependence upon luminance would be so prominent that the more important effects of chromaticity variations would be concealed in the X, Y, Z representation. Tabulated values of G_{ik} are adequate for all applications of the data to color specifications in terms of X, Y, Z. Ellipsoids will be shown only for the x, y, l representations.

Silberstein has given formulas for determining the directions and lengths of the principal axis of the ellipsoids defined by

$$g_{11},\ g_{12},\ g_{22},\ g_{23},\ g_{33}\text{ and }g_{13}.$$[8]

The procedure he suggested has been simplified for numerical computation, as follows:

Let $g_0=|g_{ik}|$ be the determinant of g_{ik}.
Let $\mathbf{G}_{ik}$ be the minor or co-factor of g_{ik} in that determinant.
Let $g_1=\mathbf{G}_{11}+\mathbf{G}_{22}+\mathbf{G}_{33}$, and $g_2=g_{11}+g_{22}+g_{33}$.

Then, one root, σ_1, of the cubic equation formed with those coefficients,

$$\sigma^3-g_2\sigma^2+g_1\sigma-g_0=0$$

can be found most conveniently by Horner's approximation method. That method also furnishes a quadratic equation such that when σ_1 is added to its roots the remaining roots, σ_2 and σ_3 of the original cubic equation are obtained. Then the lengths of the principal axis of the discrimination ellipsoid are

$$a_1=\sigma_1^{-\frac{1}{2}},\quad a_2=\sigma_2^{-\frac{1}{2}},\quad\text{and}\quad a_3=\sigma_3^{-\frac{1}{2}}.$$

The ratios of the direction cosines of the first axis are

[4] D. L. MacAdam, J. Opt. Soc. Am. **33**, 18–26 (1943).
[5] Hecht, Peskin, and Patt, J. Gen. Physiol. **22**, 7–19 (1938).
[6] D. L. MacAdam, J. Opt. Soc. Am. **32**, 247–274 (1942).
[7] A. W. Wundheiler, J. Opt. Soc. Am. **36**, 288–291 (1946).
[8] L. Silberstein, J. Opt. Soc. Am. **36**, 464–468 (1946).

given by

$$c_1:c_2:c_3=[(g_{22}-\sigma_1)(g_{33}-\sigma_1)-g_{23}^2]:$$
$$[g_{13}g_{23}-g_{12}(g_{33}-\sigma_1)]:[g_{12}g_{23}-g_{13}(g_{22}-\sigma_1)].$$

The ratios of the direction cosines of the second axis are given by substituting σ_2 in place of σ_1 in that formula. Likewise, the ratios of the direction cosines of the third axis are given by substituting σ_3 in place of σ_1. The longitude, ϕ, of any axis, measured from the y, log$Y$ plane, is given by $\tan\phi=c_1/c_2$. The colatitude, θ, of that axis measured from the log$Y$ axis is then given by $\tan\theta=(c_2/c_3)(1+c_1^2/c_2^2)^{\frac{1}{2}}$.

Although they are necessary for the construction of models representing the results, the principal axes of the discrimination ellipsoids have no intrinsic significance. Their orientations and lengths and, more fundamentally, the series of colors corresponding to each one are changed by every change of scale in the luminance dimension. A more informative specification of color discrimination, not dependent upon the arbitrary choice of a luminance scale, is given by the constant-luminance cross section through the center and by the chromaticity and luminance displacement (from the center) of the point of maximum luminance on the ellipsoid. All constant-luminance cross sections of any one ellipsoid have the same ratio of major to minor axis and the same orientation of those axes as the cross section through the center. The centers of such cross sections are displaced from the center of the central sections by amounts proportional to ΔY, as shown in Fig. 1. This displacement is an intrinsic and useful measure of the asymmetry of the discrimination surface with respect to luminance variations.

The maximum value of Δl represented by any point on the ellipsoid is given by $\Delta l_m=(\mathbf{G}_{33}/g_0)^{\frac{1}{2}}$, where the symbols are as defined previously. The corresponding luminance difference is given by $\Delta Y_m=11.53\Delta l_m$. The chromaticity displacements of this point, from the center of the discrimination ellipsoid, are given by

$$\Delta x_m=\mathbf{G}_{13}/\Delta l_m\cdot g_0 \quad \text{and} \quad \Delta y_m=\mathbf{G}_{23}/\Delta l_m\cdot g_0.$$

The displacement of the center of any cross section at constant luminance is proportional to ΔY. The length of the radius in any specific direction, including the major and minor axes, is proportional to

$$[1-(\Delta Y/\Delta Y_m)^2]^{\frac{1}{2}}.$$

III. APPARATUS

The colorimeter used in this investigation exhibits a two-degree matching field in a dark surround, which is observed monocularly with the natural pupil. The observer makes a color-match by adjusting three control knobs, each of which controls one of the three primaries in the additive mixture. When matched, the two halves of the field have the same spectral-energy distribution. The intensities of the three primaries are varied by utilizing the inverse-square variation of illuminance with distance. The colorimeter itself is shown in Figs. 2, 3, and 4. The light paths in the instrument are shown by dashed lines. The light sources are three frosted lamps (110-volt lamps operating at 117 volts), L_1, L_2, L_3, in the balancing side and one of the same type, L_4, in

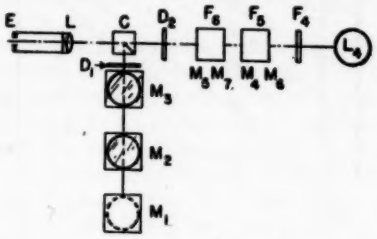

FIG. 3. End view of colorimeter.

the comparison beam. The lamps are run above the rated voltage in order to obtain a color temperature of 2850°K. Each of the lamps which illuminate the balancing field, 1, moves in a slide, the lamps being oriented so that the maximum of their polar light distribution was directed along the optic axis. This was done by attaching a drop of solder to the bottom contact of the lamp and filing this drop until the lamp was in the correct position when screwed tightly into the socket.

The contribution to the luminance of the diffuser, D_1, due to each of the three beams, is varied by changing the lamp distance. The variation of intensity is proportional to the inverse square of the optical distance of the lamp from D_1. The lamps are mounted in carriages which slide in guides. The positions of the lamps are indicated by meter-sticks attached to the lamp carriages. These meter-sticks also serve as the means of moving the lamps. To the lower edge of each is attached a rack running almost the full length of the scale. These racks rest on pinions which, in turn, are connected directly to the control knobs.

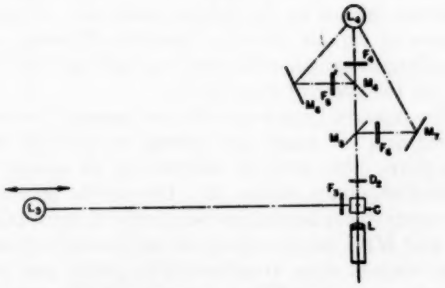

FIG. 4. Top view of colorimeter.

Rotation of the control knobs moves the meter-sticks and changes the lamp distances. A reading indicative of the lamp distance is shown on each meter-stick by a line cut in a Lucite plate mounted directly above the control knobs.

The three lamps move in separate compartments.

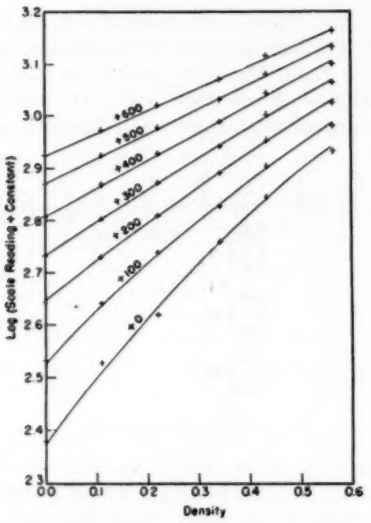

FIG. 5. Calibration curves for various assumed values of the optical distance from the scale index to the diffuser.

Thus, only light from one source passes through each filter. The filters, F_1, F_2, F_3, providing the primary colors, are mounted in the end of the lamp compartments. In order to extend the range of luminance variation to lower values than are obtainable by moving the lamp, calibrated neutral-density filters are superimposed upon the primary filters. The densities of these are chosen so that the interchange of successive filters of the set changes the lamp position required for a given intensity from one end to the other of the range of movement. Thus, a continuous variation over a wide range can be obtained by the proper choice of neutral filters.

The lamp compartments are painted dull black to eliminate reflections. Such reflections would cause deviations from the inverse-square law, because the reflected images of the lamps would act as separate sources at various distances from the diffusing screen. Diaphragms are placed around the light path to reduce further the effect of stray light.

The primary beams are reflected upward by mirrors placed at a 45° angle just beyond the primary filters. The lower (red) beam is reflected by an opaque first-surface aluminum mirror, M_1. The middle (green) and the upper (blue) beams are reflected by dichroic mirrors, M_2 and M_3. Complex layers on one surface of each of these mirrors cause them partly to reflect and partly to transmit light incident upon them, with very little loss by absorption. This is an effect of interference; consequently, the reflectance and transmittance vary considerably with wave-length. The mirrors have a relatively high reflectance in the blue and a relatively high transmittance in the red.

As can be seen in Fig. 2, the red beam passes through two of these mirrors, the green beam is reflected once at M_2, and transmitted once through M_3. The blue beam undergoes but one reflection, at M_3. This arrangement utilizes the properties of the mirrors to their best advantage.

The mirrors are adjusted so that the optic axes of the primary beams are collinear, thus combining the three beams. This combined beam then strikes the small diffusing screen, D_1.

The light from this diffusing screen fills the lower half, 1, of the field of Lummer-Brodhun cube, C, and is viewed by a low powered telescope, L, shown in the end view, Fig. 3. The field of the telescope subtends an angle of two degrees at the eye. The telescope was fitted with a sponge-rubber shield of the type used in aircraft drift-sights. This, in turn, was surrounded by an opaque shield which is not shown in the diagram. This shield protected the observer's eyes from stray light.

The comparison beam is shown in the top view of the apparatus, Fig. 4. A single lamp, L_4, is used in this case, since no variation in lamp distance is required. The filters are placed in such positions as to simulate the conditions in the variable half of the field. The beam from the red filter, F_4, passes through two dichroic mirrors, M_4, and M_5. The beam from the green filter, F_5, is once reflected (from M_4) and once transmitted (M_5), while the beam from the blue filter, F_6, is once reflected (from M_5). The mirrors, M_6 and M_7, are first-surface aluminum mirrors placed so as to reflect the light from the lamp through the blue and green filters. Since it is necessary to vary the color of the comparison half of the field only about once each week and then only in discrete steps, all changes are made by changing neutral filters superimposed on the red, green, and blue filters. In this manner, a close approximation to any desired point within the chromaticity triangle defined by the primaries can be produced.

After reflection from the mirrors, the three beams are incident upon a diffusing screen, D_2, identical with the one, D_1, in the balancing side of the colorimeter. The diffuser, D_2, is placed at the same distance from the photometric cube as the diffuser, D_1, and is viewed in the top half, 2, of the field. In this way the field is physically as well as psychophysically matched. This reduces the effects of individual differences of color vision to a minimum.

As first built, the lamps were moved vertically and the comparison system consisted merely of a light source and a holder for a single filter. After a few months of use, some disadvantages were recognized, the correction of which resulted in the present apparatus. A black deposit which formed inside the tops of the lamps introduced the effect of a gradually increasing neutral density in the beams. The heat from the lamps also set up convection currents in the vertical lamp compartments which carried dust upward and deposited it upon the primary filters. The cleaning which was frequently necessary usually disturbed calibration. A more elaborate comparison system was adopted in order to provide

a physical match and to permit more flexibility in the choice of colors to be matched.

IV. CALIBRATION

The distance, L, between each lamp and the diffusing screen is measured by the reading on the meter-stick plus the constant distance from the indicator line to the diffuser. This constant is obviously different for each of the three primaries. The length of the light path from the indicator line to the diffusing screen is 255 mm for the blue beam, 350 mm for the green, and 455 mm for the red. If spurious reflections and stray light are adequately eliminated, then the illuminance is given by $E=k/(S+C)^2=k/L^2$, where E is the illuminance of the diffuser, S is the distance from the lamp to the indicator line, and C is the length of the light path from the indicator line to the diffuser. The constant, k, is proportional to the luminous intensity of the lamp.

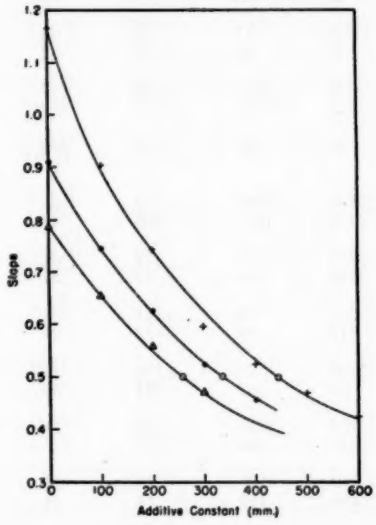

FIG. 6. Slopes of calibration curves for various assumed values of optical distances from the scale indices to the diffuser.

The calibration formula just given can be expressed as $\log E=\log k_1-2\log(S+C)$. When two of the primary beams are blocked by an opaque card, movement of the control knob for the third beam provides a method of obtaining a photometric match. Calibrated neutral-density filters were placed in the fixed half of the field so that the luminance of the diffuser, D_2, was $B=B_0T$, where T is the transmittance of each neutral filter. This expression can be written

$$\log B=\log B_0+\log T=\log B_0-D,$$

where D is the optical density of the filter.

The luminance, B, of the diffuser, D_1, is proportional to its illuminance, E; hence, $\log B=\log E+\log k_2$. When a photometric match is obtained, the two values of B (and $\log B$) are equal, therefore,

$$\log E+\log k_2=\log B_0-D.$$

Substituting the value of $\log E$ indicated above, we get the calibration equation, $D=2\log(S+C)+k_3$, where k_3 is a constant of no importance in the calibration.

Photometric matches were made with a set of calibrated neutral filters in the fixed half, 2, of the field. This was done for each of the primary beams in turn. A curve of $\log(S+C)$ plotted against neutral density was drawn for each of the primaries. A number of curves were plotted with slightly different values of C. These curves are shown in Fig. 5. The slopes of successive curves of this family decrease gradually with increase of C. It may be noted, too, that for constants that are far too small, the curves depart from linearity. The slope of each curve was measured and a curve showing this slope as a function of C is plotted in Fig. 6. From this curve the value of C that gives a slope of one half was determined. A slope of one half is required since $\log(S+C)$ is plotted against the independent variable, D. The constant, C, was found to be equal to the distance measured along the light path from the scale index to the diffuser, within the limits of accuracy of the measurements. This verifies the inverse-square law for use in this instrument. Before diaphragms were inserted around the beams, no such agreement could be obtained, because unwanted reflections and stray light caused a considerable difference between the measured

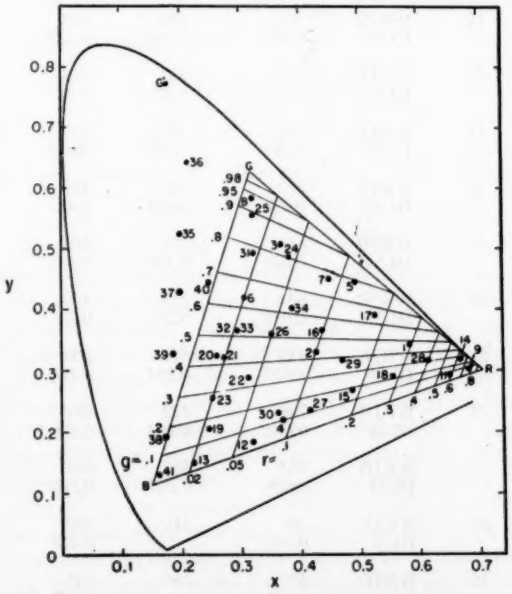

FIG. 7. Chromaticity diagram showing primaries used in colorimeter, filters used in calibration, coordinate system based on one calibration, and colors around which discrimination was investigated.

TABLE I. Experimental means (in millimeters).

Color	Obs.	$S_r(T_r)$	$S_g(T_g)$	$S_b(T_b)$	$(\Delta S_r)^2$	$(\Delta S_g)^2$	$(\Delta S_b)^2$	$\Delta S_g \cdot \Delta S_b$	$\Delta S_r \cdot \Delta S_b$	$\Delta S_r \cdot \Delta S_g$
1	WRJB	316	291	537	49	79	204	29	7	−2
	DLM		*0.437*	*0.416*	89	109	112	12	11	24
2	WRJB	549	390	436	59	36	67	6	24	18
	DLM				98	48	52	15	16	29
3	WRJB	634	290	672	86	18	282	−2	30	14
	DLM	*0.306*		*0.416*	132	55	372	30	11	26
4	WRJB	678	580	351	62	753	35	−37	26	−7
	DLM		*0.194*		195	829	55	28	72	38
5	WRJB	367	498	475	37	29	400	13	20	18
	DLM	*0.306*		*0.091*	53	46	400	8	6	29
6	WRJB	408	471	458	67	23	71	6	43	5
	DLM	*0.098*		*0.416*	58	37	48	8	−2	13
7	WRJB	370	270	599	34	18	168	9	17	16
	DLM	*0.436*		*0.416*	27	10	111	5	−8	2
8	WRJB	515	389	562	2480	20	688	−21	580	−84
		0.011		*0.089*						
9	WRJB	641	556	652	96	3165	1100	224	65	132
	DLM		*0.100*	*0.089*	204	5320	1675	840	44	−23
11	WRJB	633	564	506	70	3445	169	228	32	127
	DLM		*0.100*	*0.270*	134	2785	166	27	57	−92
12	WRJB	802	377	820	104	827	99	−161	33	7
	DLM	*0.092*	*0.025*		56	317	88	−32	25	3
13	WRJB	690	646	842	96	1260	71	−134	46	−112
	DLM	*0.027*	*0.025*		49	676	67	−104	30	−35
14	WRJB	725	520	396	80	973	1280	60	14	80
	DLM	*0.436*	*0.100*	*0.020*	151	625	784	133	7	−68
15	WRJB	592	589	623	78	262	127	−20	34	57
	DLM	*0.092*	*0.446*	*0.270*	52	476	113	44	24	16
16	WRJB	528	582	615	50	31	125	6	13	20
	DLM	*0.092*	*0.446*	*0.270*	48	46	62	0	−11	21
17	WRJB	792	455	459	117	37	246	12	17	30
	DLM	*0.291*	*0.446*	*0.089*	90	50	121	32	−6	24
18	WRJB	502	702	447	72	384	98	33	22	68
	DLM	*0.436*	*0.448*	*0.428*	77	361	42	−35	11	7
19	WRJB	486	419	464	96	107	77	−26	51	−10
	DLM	*0.026*	*0.100*	*0.428*	81	61	52	−8	36	9
20	WRJB	634	315	450	404	25	64	1	69	−27
	DLM	*0.026*	*0.293*	*0.428*	318	23	36	12	−7	−3
21	WRJB	516	312	300	131	29	28	1	18	−8
	DLM	*0.026*	*0.293*	*0.270*	119	19	14	5	1	3
22	WRJB	798	422	297	78	35	40	1	20	9
	DLM	*0.092*	*0.293*	*0.270*	70	21	18	1	4	16
23	WRJB	552	556	302	101	49	34	0	30	−10
	DLM	*0.026*	*0.293*	*0.270*	114	58	34	11	20	15
24	WRJB	397	241	640	61	12	139	−1	24	8
	DLM	*0.092*		*0.428*	59	12	169	5	16	6
25	WRJB	499	256	308	1010	15	265	−1	196	−21
	DLM	*0.026*		*0.089*	820	6	204	2	37	3
26	WRJB	772	364	516	77	13	56	1	0	7
	DLM	*0.291*			38	8	45	5	−2	3

TABLE I.—*Continued.*

Color	Obs.	$S_r(T_r)$	$S_g(T_g)$	$S_b(T_b)$	$(\Delta S_r)^2$	$(\Delta S_g)^2$	$(\Delta S_b)^2$	$\Delta S_g \cdot \Delta S_b$	$\Delta S_r \cdot \Delta S_b$	$\Delta S_r \cdot \Delta S_g$
27	WRJB	641	615	408	61	813	47	−81	21	50
	DLM	*0.436*	*0.293*		83	796	81	−78	42	−6
28	WRJB	744	544	694	58	241	216	65	26	44
	DLM		*0.446*	*0.428*	47	431	150	−94	−17	−15
29	WRJB	300	352	394	22	21	25	0	6	9
	DLM	*0.436*			23	18	39	2	5	−1
30	WRJB	684	572	411	51	114	37	−20	19	16
	DLM	*0.291*	*0.446*		45	103	50	−23	9	−6
31	WRJB	378	350	458	300	10	96	−4	76	−17
	DLM	*0.026*		*0.270*	205	6	92	2	0	−11
32	WRJB	595	355	514	134	13	34	−1	19	−10
	DLM	*0.092*			71	9	37	1	−6	2
33	WRJB	544	361	477	201	28	153	−27	47	−21
	DLM	*0.005*	*0.055*	*0.055*	129	18	98	−13	37	−2
34	WRJB	719	310	652	36	13	79	2	18	9
	DLM	*0.291*			57	9	53	5	8	2
35	WRJB	611	522	432	1752	38	170	0	153	−86
	DLM	*0.00756*		*0.089*	739	25	124	4	47	−13
36	WRJB	604	522	359	1247	19	1088	36	512	−65
	DLM	*0.00756*		*0.02405*	945	24	392	−16	−19	−37
37	WRJB	356	522	601	426	20	65	−1	84	−61
	DLM	*0.00756*		*0.270*	192	27	105	−8	9	16
38	WRJB	311	565	398	486	77	59	−137	89	−35
	DLM	*0.00756*			307	109	82	−6	75	−19
39	WRJB	347	543	560	311	28	73	−10	70	−28
	DLM	*0.00756*		*0.428*	364	36	53	21	13	−42

distance and the value of C necessary to yield a slope of one half.

The shape of the curve in Fig. 5, for the optimum value of C, shows that the deviation from linearity due to the finite size of the source is not appreciable. This would be expected from the dimensions of the apparatus. The error as calculated for the extreme positions of the lamps is less than one percent.

The contribution of each of three primaries to the color of the diffusing screen, D_1, can be written as

$$R=10^6T_r/(S_r+C_r)^2,\quad G=10^6T_g/(S_g+C_g)^2,$$
$$B=10^6T_b/(S_b+C_b)_2.$$

The values of $T_i(i=r,\ g,\ b)$ are the transmittances of the neutral filters placed in the primary beams. The values of S_i are the scale readings expressed in millimeters and those of C_i are the appropriate constants determined in the manner just described. As usual in colorimetry, the units in which the primaries, R, G, B, are measured are arbitrary. The formulas adopted above yield values of convenient magnitude. In order to express the results in terms of the standard ICI primaries rather than the instrument primaries, it is necessary to determine the numerical coefficients of the following transformation formulas:

$$R=a_{11}X+a_{12}Y+a_{12}Z$$
$$G=a_{21}X+a_{22}Y+a_{23}Z$$
$$B=a_{31}X+a_{32}Y+a_{33}Z,$$

or more concisely, in matrix notation,

$$\{R, G, B\}=a_{ik}(X, Y, Z).$$

Since the selective characteristics of the partially transparent mirrors influence the colors of the primaries, and since this effect cannot be evaluated directly with a spectrophotometer, the desired transformation cannot be deduced from the spectrophotometric curves of the primary filters. Therefore, the following indirect method was used: The mirrors and the filters were removed from the comparison beam and were replaced by a single color-filter. The amounts of the primaries, R, G, and B, necessary to match this filter were determined from the averages of ten matchings. The tristimulus specification of each filter used in a comparison field was determined from its spectrophotometric curve. The value of Y in each case was the luminous transmittance of the filter, and the values of X and Z were consistent with that convention.

It is clear from the equations, $\{R, G, B\}=a_{ik}(X, Y, Z)$,

TABLE II. Standard deviations and correlations of instrument primaries.

Color	Obs.	R	G	B	σ_r	σ_g	σ_b	ρ_{gb}	ρ_{rb}	ρ_{gr}
1	WRJB	1.682	1.060	0.663	0.0305	0.0295	0.0239	0.23	0.09	−0.04
	DLM	1.818	1.108	0.720	0.0464	0.0368	0.0201	0.10	0.10	0.24
2	WRJB	0.962	1.820	2.085	0.0147	0.0295	0.0495	0.12	0.38	0.38
	DLM	1.013	1.882	2.201	0.0202	0.0357	0.0471	0.31	0.23	0.43
3	WRJB	0.258	2.441	0.485	0.0044	0.0320	0.0176	−0.03	0.19	0.35
	DLM	0.295	2.472	0.517	0.0067	0.0575	0.0222	0.21	0.05	0.31
4	WRJB	0.779	0.228	2.723	0.0108	0.0133	0.0530	−0.23	0.56	−0.03
	DLM	0.775	0.218	2.687	0.0181	0.0110	0.0653	0.13	0.70	0.09
5	WRJB	0.455	1.391	0.171	0.0068	0.0177	0.0094	0.12	0.16	0.56
	DLM	0.465	1.397	0.185	0.0084	0.0224	0.0105	0.06	0.04	0.58
6	WRJB	0.132	1.481	0.818	0.0025	0.0173	0.0193	0.14	0.62	0.12
	DLM	0.134	1.480	0.810	0.0024	0.0219	0.0156	0.18	−0.03	0.29
7	WRJB	0.641	2.600	0.571	0.0092	0.0352	0.0173	0.16	0.22	0.65
	DLM	0.671	2.660	0.588	0.0083	0.0278	0.0144	0.15	−0.15	0.13
8	WRJB	0.0117	1.828	0.134	0.0012	0.0219	0.0086	−0.18	0.44	−0.38
9	WRJB	0.839	0.122	0.108	0.0150	0.0152	0.0079	0.12	0.20	0.24
	DLM	0.852	0.119	0.105	0.0224	0.0189	0.0093	0.28	0.05	−0.02
11	WRJB	0.849	0.120	0.466	0.0131	0.0154	0.0159	0.28	0.29	0.27
	DLM	0.849	0.124	0.455	0.0181	0.0146	0.0156	0.04	0.40	−0.15
12	WRJB	0.060	0.048	0.869	0.0009	0.0038	0.0162	−0.59	0.34	0.01
	DLM	0.058	0.056	0.870	0.0007	0.0030	0.0152	−0.19	0.36	0.02
13	WRJB	0.021	0.025	0.833	0.0004	0.0018	0.0128	−0.45	0.56	−0.32
	DLM	0.021	0.027	0.850	0.0003	0.0014	0.0128	−0.49	0.52	−0.19
14	WRJB	0.314	0.132	0.047	0.0048	0.0094	0.0052	0.05	0.04	0.29
	DLM	0.312	0.126	0.052	0.0065	0.0070	0.0047	0.19	0.01	−0.22
15	WRJB	0.100	0.103	0.353	0.0019	0.0034	0.0091	−0.11	0.34	0.40
	DLM	0.100	0.103	0.348	0.0018	0.0045	0.0084	0.19	0.33	0.10
16	WRJB	0.095	0.515	0.358	0.0014	0.0062	0.0092	0.09	0.16	0.49
	DLM	0.096	0.508	0.370	0.0013	0.0074	0.0068	0.00	−0.20	0.45
17	WRJB	0.188	0.690	0.175	0.0033	0.0105	0.0077	0.13	0.10	0.45
	DLM	0.184	0.683	0.186	0.0028	0.0120	0.0059	0.41	−0.06	0.35
18	WRJB	0.476	0.405	0.870	0.0085	0.0151	0.0245	0.17	0.26	0.41
	DLM	0.471	0.384	0.868	0.0086	0.0135	0.0164	−0.28	0.19	0.04
19	WRJB	0.029	0.170	0.830	0.0006	0.0045	0.0203	−0.26	0.59	−0.09
	DLM	0.029	0.173	0.850	0.0006	0.0036	0.0173	−0.16	0.55	0.14
20	WRJB	0.022	0.663	0.863	0.0008	0.0100	0.0196	0.02	0.43	−0.27
	DLM	0.021	0.668	0.864	0.0007	0.0095	0.0143	0.42	−0.08	−0.03
21	WRJB	0.028	0.671	0.877	0.0006	0.0110	0.0164	0.04	0.30	−0.13
	DLM	0.027	0.675	0.877	0.0006	0.0090	0.0117	0.33	0.03	0.07
22	WRJB	0.059	0.492	0.889	0.0008	0.0075	0.0203	0.04	0.37	0.18
	DLM	0.058	0.498	0.893	0.0008	0.0060	0.0140	0.04	0.12	0.42
23	WRJB	0.026	0.357	0.871	0.0005	0.0055	0.0181	0.00	0.51	−0.13
	DLM	0.025	0.360	0.877	0.0005	0.0061	0.0182	0.25	0.32	0.18
24	WRJB	0.127	2.86	0.535	0.0023	0.0339	0.0141	−0.02	0.26	0.30
	DLM	0.133	2.95	0.554	0.0025	0.0355	0.0164	0.12	0.16	0.21
25	WRJB	0.029	2.72	0.282	0.0019	0.0350	0.0163	−0.02	0.38	−0.17
	DLM	0.027	2.76	0.287	0.0016	0.0220	0.0147	0.07	0.09	0.05
26	WRJB	0.194	1.965	1.68	0.0028	0.0198	0.0326	0.01	0.01	0.22
	DLM	0.188	2.035	1.66	0.0019	0.0168	0.0286	0.26	−0.04	0.17

TABLE II.—*Continued.*

Color	Obs.	R	G	B	σ_r	σ_g	σ_b	ρ_{gb}	ρ_{rb}	ρ_{gr}
27	WRJB	0.365	0.321	2.28	0.0052	0.0190	0.0474	−0.41	0.40	0.23
	DLM	0.360	0.320	2.24	0.0060	0.0189	0.0604	−0.31	0.51	−0.02
28	WRJB	0.700	0.560	0.475	0.0089	0.0193	0.0147	0.29	0.22	0.38
	DLM	0.715	0.526	0.460	0.0083	0.0236	0.0115	−0.36	−0.20	−0.11
29	WRJB	0.762	2.035	2.36	0.0095	0.0266	0.0364	0.00	0.26	0.42
	DLM	0.780	2.05	2.39	0.0100	0.0246	0.0465	0.09	0.18	−0.05
30	WRJB	0.226	0.526	2.25	0.0028	0.0121	0.0401	−0.30	0.43	0.21
	DLM	0.226	0.523	2.28	0.0027	0.0115	0.0490	−0.31	0.19	−0.08
31	WRJB	0.038	2.05	0.531	0.0016	0.0187	0.0146	−0.12	0.45	−0.30
	DLM	0.032	2.10	0.510	0.0010	0.0152	0.0134	0.08	0.00	−0.32
32	WRJB	0.083	2.015	1.695	0.0017	0.0206	0.0256	−0.05	0.31	−0.26
	DLM	0.080	2.05	1.70	0.0013	0.0176	0.0194	0.04	−0.12	0.08
33	WRJB	0.005	0.109	0.103	0.0001	0.0016	0.0035	−0.41	0.27	−0.28
	DLM	0.005	0.112	0.092	0.0001	0.0013	0.0024	−0.30	0.29	−0.04
34	WRJB	0.212	2.30	1.22	0.0022	0.0251	0.0240	0.05	0.34	0.41
	DLM	0.225	2.29	1.20	0.0028	0.0209	0.0197	0.23	0.15	0.09
35	WRJB	0.00680	1.319	0.188	0.000533	0.0186	0.00715	0.00	0.28	−0.16
	DLM	0.00640	1.310	0.192	0.000320	0.0150	0.00628	0.07	0.16	−0.04
36	WRJB	0.00676	1.319	0.0640	0.000452	0.01319	0.00688	0.19	0.44	−0.42
	DLM	0.00693	1.319	0.0727	0.000408	0.01480	0.00499	−0.17	−0.31	−0.24
37	WRJB	0.01146	1.319	0.368	0.000583	0.01348	0.00694	−0.03	0.50	−0.66
	DLM	0.01140	1.315	0.362	0.000540	0.01570	0.00860	0.31	0.05	−0.08
38	WRJB	0.0128	1.200	2.34	0.000738	0.0231	0.0551	−0.62	0.52	−0.51
	DLM	0.0128	1.200	2.34	0.000587	0.0274	0.0651	−0.03	0.47	−0.10
39	WRJB	0.0118	1.256	0.646	0.000518	0.0149	0.0135	−0.30	0.47	−0.23
	DLM	0.0118	1.256	0.646	0.000550	0.0168	0.0114	0.49	0.09	−0.37

that only three colors need be matched to determine values of a_{ik}. Indeed this was done at first, but the resulting coefficients were not considered sufficiently accurate. To improve the accuracy of determination, more than three colors were matched and a least-squares method was used to determine the calibration. The calibrations which have been used throughout the majority of these determinations were based on the matching of seven color filters. Six of these gave colors near the edge of the region covered by the instrument primaries while the seventh was a white in the center of that region. The filters used are shown in Fig. 7. They are indicated by the numbers 1, 4, 5, 8, 34, 40, 41.

The inverse of the calibration equations indicated above are also useful. These are written

$$X=A_{11}R+A_{12}G+A_{13}B$$
$$Y=A_{21}R+A_{22}G+A_{23}B$$
$$Z=A_{31}R+A_{32}G+A_{33}B.$$

These equations can be obtained from the original calibration equations by noting that $A_{ik}=a_{ik}'/|a_{ik}|$, where a_{ik}' is the minor of a_{ik} in the matrix formed by the coefficients of the original calibration equations, and $|a_{ik}|'$ is the determinant of this matrix. To obtain the values of Y in foot-lamberts and to convert the values of X and Z consistent with this convention, it is necessary to multiply each coefficient, A_{ik}, by a common factor. This factor was obtained by adjusting the colorimeter to exhibit a yellow-white color and measuring the luminance of the field with a Macbeth illuminometer. The factor by which all of the A_{ik}'s must be multiplied is Y_0/Y_c, where Y_c is the value of the field luminance calculated from the scale readings and the inverse transformation equations indicated, and Y_0 is the measured luminance of the field. The original calibration coefficients, a_{ik}, were also multiplied by the reciprocal of this ratio (i.e., Y_c/Y_0) to obtain calibration equations consistent with Y in foot-lamberts.

The inverse calibration equations were used to determine the location of points on the ICI chromaticity diagram. The grid work shown in Fig. 7 was derived from these equations. Its parameters are

$$r=R/(R+G+B) \quad \text{and} \quad g=G/(R+G+B).$$

It is to be noted that the calibration places one of the primaries (the red) just outside the spectrum locus. This primary is certainly very saturated, the filter employed having a purity of 100 percent. Therefore, the point representing it should be on the spectrum locus. However, the least-squares method of calibration

TABLE III. Discrimination coefficients for instrument primaries.

Color	Obs.	c_{11}	c_{12}	c_{22}	c_{23}	c_{33}	c_{13}
1	WRJB	545	36	610	−177	936	−72
	DLM	248	−73	395	−56	1,260	−44
2	WRJB	3,120	−528	672	11	239	−314
	DLM	1,523	−343	512	−86	253	−70
3	WRJB	30,800	−1,510	564	102	1,695	−1,550
	DLM	12,350	−450	175	−88	1,060	57
4	WRJB	6,335	−536	3,030	236	278	−753
	DLM	2,993	5	4,213	−93	232	−580
5	WRJB	15,980	−3,365	2,320	−135	5,810	−1,088
	DLM	10,680	−2,325	1,505	−131	4,550	−45
6	WRJB	130,000	−636	1,708	−163	2,195	−10,375
	DLM	95,500	−3,185	1,182	−315	2,145	1,242
7	WRJB	10,460	−1,730	701	−25	1,758	−664
	DLM	7,600	−353	679	−228	2,532	761
8	WRJB	487,000	8,300	1,217	50	8,400	−2,613
9	WRJB	2,450	−528	2,310	−333	8,400	−808
	DLM	996	44	1,522	−872	6,300	−146
11	WRJB	3,320	−580	2,380	−333	2,260	−630
	DLM	1,878	386	2,418	−270	2,467	−883
12	WRJB	707,000	−56,000	57,500	9,050	3,575	−21,550
	DLM	1,195,000	−25,500	59,000	2,620	2,600	−20,650
13	WRJB	5,800,000	97,500	195,000	10,870	5,090	−84,350
	DLM	10,650,000	−162,500	326,000	19,450	5,330	−118,800
14	WRJB	23,700	−3,490	6,165	−430	18,520	560
	DLM	12,450	2,670	11,150	−3,200	23,550	933
15	WRJB	211,000	−51,000	56,000	5,840	7,400	−16,600
	DLM	175,000	−2,675	25,900	−2,442	8,240	−12,080
16	WRJB	355,500	−38,000	17,260	−141	6,070	−6,210
	DLM	364,000	−29,900	11,680	−1,172	11,200	14,350
17	WRJB	59,100	−8,180	5,790	−676	8,600	−1,075
	DLM	77,600	−8,140	5,025	−4,440	18,425	9,010
18	WRJB	8,770	−1,858	2,660	−112	898	−592
	DLM	7,090	−456	3,005	741	2,108	−812
19	WRJB	2,142,000	−19,250	26,700	1,878	1,995	−38,600
	DLM	2,568,000	−92,000	43,000	3,035	2,610	−48,000
20	WRJB	1,020,000	23,320	5,590	−472	1,640	−18,430
	DLM	1,025,000	−376	6,720	−1,880	2,980	4,100
21	WRJB	1,342,000	11,300	4,280	−247	2,044	−16,140
	DLM	1,412,000	−6,340	6,970	−1,760	4,100	−1,003
22	WRJB	880,000	15,930	9,150	102	1,410	−13,000
	DLM	1,046,000	−55,800	17,100	141	2,590	−5,860
23	WRJB	2,450,000	30,700	16,820	−456	2,075	−36,650
	DLM	2,025,000	−18,750	14,675	−1,048	1,758	−17,225
24	WRJB	111,000	−2,313	483	122	2,728	−4,850
	DLM	85,500	−1,165	418	−80	1,920	−1,780
25	WRJB	165,200	1,460	422	−47	2,200	−7,295
	DLM	197,000	−640	1,034	−104	2,340	−1,850
26	WRJB	67,600	−2,090	1,340	−8	469	11
	DLM	143,500	−3,150	1,965	−310	660	867

TABLE III.—*Continued.*

Color	Obs.	c_{11}	c_{12}	c_{22}	c_{23}	c_{33}	c_{13}
27	WRJB	28,300	−3,680	2,148	513	388	−1,840
	DLM	19,300	−936	1,590	203	212	−1,070
28	WRJB	7,315	−1,072	1,620	−488	2,550	−478
	DLM	4,835	453	1,075	876	4,660	552
29	WRJB	7,165	−1,050	815	72	399	−485
	DLM	5,190	140	842	−45	241	−208
30	WRJB	91,350	−7,905	4,440	641	474	−3,475
	DLM	71,200	386	4,190	301	238	−715
31	WRJB	282,000	5,820	1,568	−38	2,945	−12,520
	DLM	558,500	11,800	2,430	−220	2,812	−1,082
32	WRJB	187,000	4,140	1,260	−35	842	−4,040
	DLM	302,500	−1,900	1,630	−74	1,350	2,495
33	WRJB	26,850,000	493,000	246,000	40,500	51,250	−207,000
	DLM	44,250,000	−192,000	310,500	54,650	106,750	−631,500
34	WRJB	146,000	−5,000	963	101	997	−4,225
	DLM	65,600	−520	1,212	−284	1,386	−1,265
35	WRJB	1,963,000	9,010	1,487	−188	10,630	−41,000
	DLM	5,020,000	5,510	2,240	−419	13,090	−41,800
36	WRJB	4,420,000	80,700	4,435	−3,945	16,620	−159,500
	DLM	3,680,000	30,580	2,610	2,085	23,900	108,800
37	WRJB	4,415,000	123,200	6,200	−5,345	17,580	−178,000
	DLM	1,735,000	6,310	2,270	−1,300	7,520	−9,050
38	WRJB	1,644,000	25,220	1,650	178	246	−6,070
	DLM	1,466,000	−685	855	−170	153	1,458
39	WRJB	2,516,000	17,730	2,500	316	3,575	−41,000
	DLM	2,155,000	38,300	3,008	−2,330	5,661	−36,950

minimizes the errors made by the observer when matching the calibration filters, at the expense of slight displacements of the points representing the primaries from their true positions. Statistically, this displacement of the red point is just as likely to be outside of the spectrum locus as inside. Discrepancies between the color-mixture characteristics of the ICI observer and those of the actual observer, such as must exist if an erroneous luminosity curve was used by the ICI, would also cause inconsistencies. The least-squares solution would minimize the effects of these inconsistencies by displacing the primaries slightly from their true positions. A recalibration based on new color-matchings and remeasured filters yielded essentially the same results. It appears that the displacement of the red primary outside the spectrum locus is unavoidable and that the least-squares method provides the best over-all calibration. The original matching-data and spectrophotometric curves of the filters are being preserved so that the calibration can be revised on the basis of improved ICI data if, and when, such are adopted.

V. EXPERIMENTAL PROCEDURE

The method adopted for the greater part of the observations was based on experience gained during an extensive trial of alternative procedures. The data from these early tests have been omitted from this paper in order to make the published results as self-consistent as possible. The routine which was finally decided upon is described in the succeeding paragraphs and has proved satisfactory.

Three-color matches were made in groups of thirty (ninety scale readings) by each observer. One group was taken in the morning and one in the afternoon. It was found that results from groups smaller than thirty were statistically unreliable while groups larger than thirty caused great fatigue to the observer. Observations on a single color continued for four days in the case of the principal observer (W.R.J.B.) and for one day for the check observer (D.L.M.). This gave a total of seven hundred and twenty scale settings for the former and one hundred and eighty for the latter. It is necessary to obtain a number of determinations sufficient to yield results that are statistically reliable. If the number of observations on each color were too great, the experiment would be prolonged unnecessarily and changes in the vision and motivation of the observer during the interval would probably counteract the statistical improvement.

After determination of the calibration equations,

TABLE IV. Discrimination coefficients for ICI primaries: X, Y, Z.

Color	Obs.	X	Y	Z	G_{11}	G_{12}	G_{22}	G_{23}	G_{13}	G_{33}
1	WRJB	9.17	5.40	1.25	119	−202	424	−115	234	48
	DLM				86	−150	294	−70	306	36
2	WRJB	7.41	5.80	4.36	424	−472	645	11	64	−44
	DLM				252	−316	469	−47	65	13
3	WRJB	3.46	4.80	1.24	2,687	−1,897	1,471	190	428	−337
	DLM				1,075	−726	544	−37	259	−13
4	WRJB	4.66	2.75	5.20	958	−1,336	2,281	56	78	−126
	DLM				756	−1,340	2,838	−149	81	−4
5	WRJB	3.70	3.35	0.47	2,192	−2,315	2,780	−158	1,419	−25
	DLM				1,493	−1,577	1,866	−191	1,095	95
6	WRJB	2.14	2.95	1.86	10,181	−6,121	4,279	983	765	−1,993
	DLM				8,228	−5,466	3,976	−16	534	−196
7	WRJB	5.74	5.80	1.34	1,232	−1,137	1,149	6	435	−81
	DLM				775	−666	768	−224	603	169
8	WRJB	2.25	4.10	0.71	2,714	−1,174	783	51	1,167	−402
9	WRJB	13.75	6.10	0.32	190	−400	876	−239	1,110	88
	DLM				114	−256	593	−317	848	136
11	WRJB	13.80	6.20	1.23	200	−411	895	−152	312	50
	DLM				154	−356	858	−142	339	40
12	WRJB	1.53	0.89	2.44	10,127	−13,732	23,539	1,335	472	−1,357
	DLM				12,053	−13,855	23,719	−18	600	−1,138
13	WRJB	0.82	0.55	2.34	43,819	−41,008	72,641	661	1,907	−6,047
	DLM				92,078	−89,833	137,848	2,876	2,418	−10,166
14	WRJB	5.20	2.47	0.12	696	−1,231	2,439	−500	2,424	230
	DLM				737	−1,678	4,054	−1,333	3,167	588
15	WRJB	1.94	1.07	1.00	6,881	−11,916	22,369	321	979	−509
	DLM				2,848	−4,629	9,676	−1,029	1,283	73
16	WRJB	3.93	3.30	1.79	4,788	−5,818	8,416	−148	870	−199
	DLM				4,229	−4,649	6,195	−622	1,418	435
17	WRJB	3.88	2.88	0.63	1,069	−1,537	2,565	−340	1,154	110
	DLM				1,202	−1,604	2,554	−1,414	2,465	786
18	WRJB	8.75	4.60	2.51	300	−536	1,048	−81	128	20
	DLM				248	−487	1,062	72	265	−52
19	WRJB	1.12	0.92	2.46	15,430	−11,380	13,366	1,198	717	−2,667
	DLM				21,940	−19,183	23,059	1,651	839	−3,155
20	WRJB	1.55	1.95	2.50	5,607	−2,535	2,408	297	486	−1,233
	DLM				6,732	−4,177	4,047	−254	526	−443
21	WRJB	1.68	2.02	2.57	7,963	−3,983	2,989	521	536	−1,450
	DLM				9,331	−5,685	4,969	−3	739	−892
22	WRJB	1.99	1.85	2.58	5,285	−3,080	3,795	243	387	−983
	DLM				9,682	−8,861	10,181	314	460	−781
23	WRJB	1.32	1.32	2.54	14,711	−8,053	8,142	817	836	−2,777
	DLM				14,007	−9,328	9,118	520	588	−1,832
24	WRJB	5.72	7.30	1.91	784	−516	426	93	394	−208
	DLM				592	−384	331	17	264	−85
25	WRJB	3.59	6.20	1.32	955	−452	314	105	357	−313
	DLM				1,288	−781	695	25	368	−173
26	WRJB	6.24	6.31	5.08	583	−525	674	−5	69	−35
	DLM				1,134	−917	1,076	−65	103	−35

TABLE IV.—*Continued.*

Color	Obs.	X	Y	Z	G_{11}	G_{12}	G_{22}	G_{23}	G_{33}	G_{13}
27	WRJB	7.38	4.19	6.18	448	−606	942	100	52	−92
	DLM				253	−345	623	29	34	−46
28	WRJB	11.92	6.18	1.38	197	−340	662	−169	351	61
	DLM				76	−139	335	86	578	−27
29	WRJB	16.04	10.74	7.12	137	−200	343	5	55	−15
	DLM				80	−136	302	−30	35	5
30	WRJB	5.70	3.66	6.56	1,144	−1,368	2,035	142	77	−175
	DLM				686	−819	1,559	−3	41	−60
31	WRJB	3.21	4.91	1.88	1,545	−700	671	167	507	−539
	DLM				3,093	−1,380	1,170	39	449	−420
32	WRJB	4.58	5.71	5.28	1,046	−507	529	62	162	−244
	DLM				2,038	−1,276	1,134	20	191	−138
33	WRJB	0.251	0.313	0.289	159,156	−88,508	102,351	13,564	10,410	−28,500
	DLM				295,418	−187,208	180,933	29,833	21,804	−54,378
34	WRJB	7.24	6.96	3.88	1,138	−858	779	120	170	−223
	DLM				500	−413	565	−66	206	−41
35	WRJB	.552	1.45	.767	9,314	−2,126	2,026	211	1,431	−2,444
	DLM				24,700	−6,196	3,847	732	2,116	−5,156
36	WRJB	.455	1.38	.313	17,152	−127	3,921	−1,203	3,380	−6,224
	DLM				16,682	−2,929	3,105	220	1,614	−1,333
37	WRJB	.759	1.574	1.403	14,690	2,600	4,715	−2,273	3,711	−6,266
	DLM				8,590	−2,590	3,123	−316	976	−1,466
38	WRJB	2.298	2.551	8.622	6,742	−506	1,593	−56	374	−1,531
	DLM				7,424	−2,134	1,472	177	251	−1,242
39	WRJB	0.973	1.673	2.439	11,670	−2,515	3,114	320	989	−2,977
	DLM				8,669	−668	3,062	−914	1,195	−2,179

several Wratten filters in turn were placed in the comparison beam. A complete set of data was taken for each filter by both observers using the procedure outlined above. The fields were not, of course, physically matched but there appeared to be no important differences between the observers. Two of these filters were used to check the calibration after several months of observations. Since the chromaticity of the field was identical for both sets of observations, a comparison of the results provided an immediate check of the calibration. The change in calibration determined in this way was very slight.

When these initial data had been collected, the routine investigation of the chromaticity diagram was begun. The mirrors and primary filters were replaced in the comparison system shown in Fig. 4 so that in all subsequent matchings a complete spectral match was necessarily obtained when the two halves of the colorimeter field were visually matched.

The summations, $\sum\Delta S_i\Delta S_k$, were evaluated for the two hundred and forty tricolor matchings. The data were totaled in groups of thirty and the totals of these groups were added to obtain the final summation. This afforded a check on each group of data. The means of the scale readings were used to compute the values of R, G, and B which, in turn, were used to locate the color on a chromaticity diagram.

It is of interest to observe that the quantities, $\sum(\Delta S_i)^2/n$, are equal to the squares of the standard deviations of the instrument readings.

Also of interest is the significance of the quantity $\sum\Delta S_i\cdot\Delta S_k/n$ when i is not equal to k. If this quantity is divided by the standard deviation of each of the respective sets of data, the coefficient of correlation between them is obtained:

$$\rho_{ik}=\sum\Delta S_i\Delta S_k/(\sum\Delta S_i^2\cdot\sum\Delta S_k^2)^{\frac{1}{2}}.$$

The error to be expected in the determination of small correlation coefficients is large; the value of the probable error given by Tuttle and Satterly[9] being $0.67(1-\rho^2)n^{-\frac{1}{2}}$, where ρ is the correlation coefficient and n is the number of observations. This error is greatest for small correlations, in the case of $n=240$ being greater than the correlation itself when the latter is 0.06 or smaller. The determination of small correlation coefficients is therefore almost impossible. Increasing the number of observations is the only conceivable remedy for this unfortunate situation. This

[9] L. Tuttle and J. Satterly, *The Theory of Measurements* (Longmans, Green and Company, London, 1925).

TABLE V. Coefficients ($\times 10^{-4}$) of discrimination in x, y, and principal axes in section for $\Delta l = 0$.

Color	Obs.	x	y	l	g_{11}	g_{12}	g_{22}	g_{23}	g_{33}	g_{13}	θ	a	b
1	WRJB	0.580	0.341	0.146	6	2	21	−4	32	−1	−9°	0.0043	0.0022
	DLM				8	6	20	−8	17	−5	−21°	0.0043	0.0021
	PGN				20	−23	62				24°	0.0032	0.0012
2	WRJB	0.422	0.330	0.152	17	−13	21	−6	42	5	40°	0.0040	0.0018
	DLM				9	−7	21	−6	30	0	25°	0.0041	0.0020
	PGN				38	−26	43				43°	0.0026	0.0012
3	WRJB	0.364	0.505	0.136	34	−11	13	−2	51	−6	67°	0.0034	0.0016
	DLM				12	−4	9	−3	18	1	58°	0.0039	0.0026
	PGN				55	−12	10				76°	0.0038	0.0013
4	WRJB	0.370	0.218	0.088	21	−20	35	−2	39	3	35°	0.0038	0.0014
	DLM				14	−17	45	5	15	−2	24°	0.0039	0.0014
	PGN				80	−50	63				50°	0.0022	0.0009
5	WRJB	0.492	0.446	0.105	21	−4	25	−3	47	2	32°	0.0024	0.0019
	DLM				14	−2	20	−5	30	0	18°	0.0028	0.0014
	PGN				52	−18	28				62°	0.0023	0.0013
6	WRJB	0.308	0.424	0.094	72	−21	13	−1	62	1	72°	0.0040	0.0011
	DLM				45	−24	26	12	54	6	56°	0.0032	0.0013
	PGN				37	−11	10				70°	0.0040	0.0016
7	WRJB	0.446	0.450	0.152	31	−11	28	−8	44	3	49°	0.0023	0.0016
	DLM				18	−1	35	−20	103	5	3°	0.0024	0.0014
	PGN				56	−17	21				68°	0.0026	0.0013
8	WRJB	0.319	0.581	0.123	23	0	10	−6	62	11	90°	0.0032	0.0021
	PGN				38	−7	4				79°	0.0062	0.0016
9	WRJB	0.682	0.302	0.157	45	35	107	−14	22	2	−24°	0.0018	0.0009
	DLM				28	32	88	−8	10	−3	−30°	0.0018	0.0011
	PGN				15	−24	72				20°	0.0018	0.0009
11	WRJB	0.650	0.292	0.158	19	−1	77	−18	30	4	1°	0.0023	0.0011
	DLM				18	5	66	−1	17	−3	−6°	0.0024	0.0012
	PGN				14	−24	74				19°	0.0043	0.0011
12	WRJB	0.314	0.183	−0.010	31	−31	44	−5	48	0	39°	0.0040	0.0012
	DLM				35	−38	69	−25	58	20	32°	0.0035	0.0010
	PGN				155	−110	100				52°	0.0027	0.0006
13	WRJB	0.221	0.148	−0.052	79	−47	77	7	52	0	46°	0.0018	0.0009
	DLM				156	−119	150	9	67	14	46°	0.0017	0.0006
	PGN				300	−110	55				69°	0.0028	0.0005
14	WRJB	0.668	0.317	0.076	16	8	41	−13	30	6	−16°	0.0027	0.0015
	DLM				17	14	57	6	22	−3	−18°	0.0028	0.0013
	PGN				17	−23	68				21°	0.0035	0.0020
15	WRJB	0.484	0.267	0.006	15	−17	36	−3	24	−1	29°	0.0043	0.0015
	DLM				7	−4	23	−8	26	2	15°	0.0038	0.0021
	PGN				36	−30	69				31°	0.0023	0.0011
16	WRJB	0.436	0.366	0.104	49	−35	66	4	105	−11	17°	0.0038	0.0016
	DLM				39	−24	89	−60	117	−14	22°	0.0018	0.0010
	PGN				36	−22	35				46°	0.0027	0.0013
17	WRJB	0.525	0.390	0.092	11	0	22	1	36	−3	0°	0.0030	0.0021
	DLM				11	8	46	−11	41	0	−12°	0.0033	0.0014
	PGN				33	−20	41				39°	0.0025	0.0013
18	WRJB	0.552	0.290	0.132	10	−9	38	−11	26	2	17°	0.0038	0.0016
	DLM				15	−2	31	−8	43	−13	16°	0.0037	0.0015
	PGN				22	−26	72				23°	0.0030	0.0011
19	WRJB	0.249	0.204	−0.007	43	−20	18	1	32	2	61°	0.0035	0.0014
	DLM				58	−33	32	5	39	−4	56°	0.0032	0.0011
	PGN				260	−105	60				67°	0.0026	0.0006
20	WRJB	0.258	0.325	0.058	31	−5	6	−2	48	5	79°	0.0044	0.0018
	DLM				29	−13	19	−8	38	7	56°	0.0030	0.0016
	PGN				68	−16	20				73°	0.0026	0.0012

TABLE V.—*Continued.*

Color	Obs.	x	y	l	g_{11}	g_{12}	g_{22}	g_{23}	g_{33}	g_{13}	θ	a	b
21	WRJB	0.268	0.322	0.061	45	−13	11	−10	52	12	66°	0.0043	0.0013
	DLM				47	−19	27	−18	66	10	59°	0.0025	0.0013
	PGN				80	−25	26				69°	0.0025	0.0011
22	WRJB	0.310	0.288	0.054	31	−14	14	−10	80	21	61°	0.0040	0.0016
	DLM				48	−34	40	−12	81	5	51°	0.0028	0.0012
	PGN				143	−85	65				57°	0.0031	0.0007
23	WRJB	0.255	0.255	0.024	57	−18	15	−2	54	13	69°	0.0036	0.0013
	DLM				49	−23	22	−5	37	11	60°	0.0034	0.0013
	PGN				148	−60	40				66°	0.0027	0.0008
24	WRJB	0.383	0.489	0.172	35	0	15	−10	61	1	90°	0.0026	0.0017
	DLM				22	−2	13	−8	57	4	78°	0.0028	0.0021
	PGN				58	−14	12				74°	0.0035	0.0013
25	WRJB	0.323	0.558	0.158	24	1	5	−2	46	3	93°	0.0045	0.0020
	DLM				24	−2	9	3	105	−6	83°	0.0034	0.0020
	PGN				41	−8	4				78°	0.0065	0.0015
26	WRJB	0.354	0.358	0.160	23	−14	21	−7	98	1	43°	0.0038	0.0016
	DLM				40	−29	42	−18	134	26	43°	0.0030	0.0012
	PGN				49	−25	28				56°	0.0030	0.0012
27	WRJB	0.416	0.236	0.128	22	−19	29	−11	32	3	40°	0.0025	0.0015
	DLM				12	−10	19	−5	28	3	35°	0.0045	0.0020
	PGN				51	−38	64				40°	0.0023	0.0010
28	WRJB	0.612	0.317	0.158	16	3	59	−19	43	4	−4°	0.0025	0.0013
	DLM				27	20	39	−28	65	−22	−37°	0.0029	0.0014
					18	−23	68				21°	0.0033	0.0012
29	WRJB	0.473	0.317	0.206	25	−13	44	−10	72	−10	27°	0.0023	0.0014
	DLM				12	−8	34	8	97	−6	18°	0.0033	0.0016
	PGN				30	−26	56				32°	0.0027	0.0012
30	WRJB	0.358	0.230	0.113	40	−34	47	−12	59	6	42°	0.0033	0.0011
	DLM				21	−22	37	−11	81	11	35°	0.0042	0.0014
	PGN				89	−54	61				52°	0.0019	0.0009
31	WRJB	0.321	0.491	0.138	31	1	7	−5	117	5	92°	0.0038	0.0018
	DLM				43	−11	14	−13	174	31	71°	0.0031	0.0015
	PGN				38	−8	8				76°	0.0041	0.0016
32	WRJB	0.294	0.367	0.151	40	−6	12	−12	122	9	78°	0.0030	0.0016
	DLM				59	−27	25	−35	166	15	61°	0.0031	0.0012
	PGN				55	−17	17				69°	0.0031	0.0013
33	WRJB	0.294	0.367	−0.101	17	−6	4	−3	72	4	69°	0.0078	0.0023
	DLM				30	−10	8	−4	83	−2	69°	0.0049	0.0017
	PGN				55	−17	17				69°	0.0031	0.0013
34	WRJB	0.385	0.400	0.168	52	−21	24	−16	88	17	62°	0.0028	0.0013
	DLM				23	−4	25	−7	119	6	38°	0.0022	0.0019
	PGN				46	−19	22				61°	0.0029	0.0013
35	WRJB	0.199	0.524	0.023	10	1	2	−1	40	0	97°	0.0072	0.0032
	DLM				28	0	2	−1	58	2	90°	0.0071	0.0019
	PGN				20	4	4				103°	0.0057	0.0022
36	WRJB	0.212	0.642	0.028	15	2	1	4	107	22	98°	0.0118	0.0026
	DLM				10	−1	2	−4	77	7	83°	0.0074	0.0032
	PGN				17	2	2				98°	0.0078	0.0024
37	WRJB	0.203	0.419	0.040	43	12	7	12	142	42	107°	0.0055	0.0015
	DLM				17	4	4	1	53	3	106°	0.0059	0.0023
	PGN				26	4	6				101°	0.0044	0.0019
38	WRJB	0.171	0.189	0.081	167	5	9	12	82	2	92°	0.0033	0.0008
	DLM				238	−4	9	4	23	6	89°	0.0033	0.0006
	PGN				135	1	30				91°	0.0018	0.0009
39	WRJB	0.191	0.329	0.044	48	−3	4	0	83	−1	86°	0.0051	0.0014
	DLM				37	5	10	4	56	16	100°	0.0033	0.0016
	PGN				40	4	10				97°	0.0033	0.0016

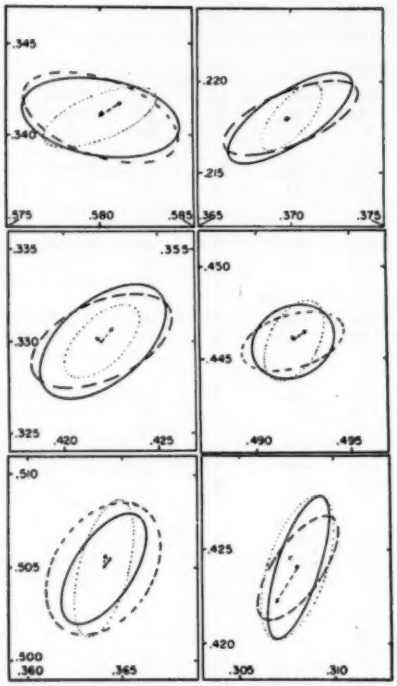

FIG. 8. Constant-luminance cross sections of discrimination ellipsoids for WRJB (solid), DLM (broken), PGN (dotted), and displacements of points of maximum luminance, for colors 1 to 3 in left column, 4 to 6 in right column.

number cannot be increased significantly greater than the number used in this investigation, for the reasons previously discussed. The present investigation entailed some 16,000 three-color matchings (48,000 scale settings) and continued for eighteen months. Despite this number of observations, the low precision of the correlation coefficients remained the limiting feature of the results.

VI. RESULTS

The averages of the color-matching data for each color are presented in Table I. Column 1 gives the serial number of the color. The second column identifies the observer. The next three columns are the average scale readings in millimeters for the instrument primaries. In cases for which neutral filters were placed in the beam, the transmittance, T_i, of the filter is shown in italics under the average scale reading. The next three columns give the mean squares of the deviations of the scale readings, $\sum\Delta S_i^2/n$, where ΔS_i is expressed in millimeters. The next three columns give the mean products of the deviations taken in pairs, $\sum\Delta S_i\cdot\Delta S_k/n$. The data for each color are presented by two rows. The uppermost row contains the data of the principal observer, based on 240 color-matches, while the lower row contains the data for the check observer, based on 60 matches. The amounts of the primaries are essentially the same for the two observers and so are given only once. All values in Table I (except T_i) have been rounded off to the nearest integer.

Table II contains results derived from the data in Table I. The first three columns following the serial numbers and observers' initials indicate the amounts of the primaries, R, G, B, present in the matching field in the arbitrary units previously defined. For computation of these quantities, the measured values of the additive constant were used, i.e., $C_r=455$ mm, $C_g=350$ mm, and $C_b=255$ mm.

The next three columns of Table II give

$$\sigma_r=(\sum\Delta R^2/n)^{\frac{1}{2}},\quad \sigma_g=(\sum\Delta G^2/n)^{\frac{1}{2}},$$

and

$$\sigma_b=(\sum\Delta B^2/n)^{\frac{1}{2}}.$$

The last three columns contain the correlation coefficients

$$\rho_{bg}=\sum\Delta G\Delta B/n\sigma_g\sigma_b,\quad \rho_{rb}=\sum\Delta R\Delta B/n\sigma_r\sigma_b,$$
$$\rho_{rg}=\sum\Delta R\Delta G/n\sigma_r\sigma_g.$$

These are given in successive rows for the two observers, as before.

Table III contains the values of the tensor, c_{ik}, for both observers. These are computed from the equation, $c_{ik}=\rho^{ik}/2\sigma_i\sigma_k$, as shown in Section II. Table III gives

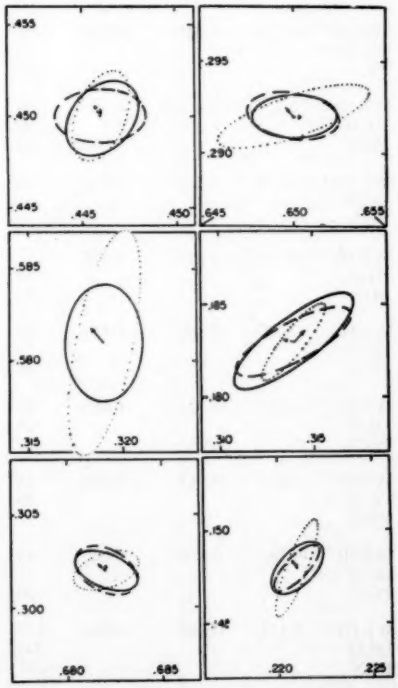

FIG. 9. Constant-luminance cross sections of discrimination ellipsoids and displacements of points of maximum luminance for colors 7 to 9 (left), 11 to 13 (right).

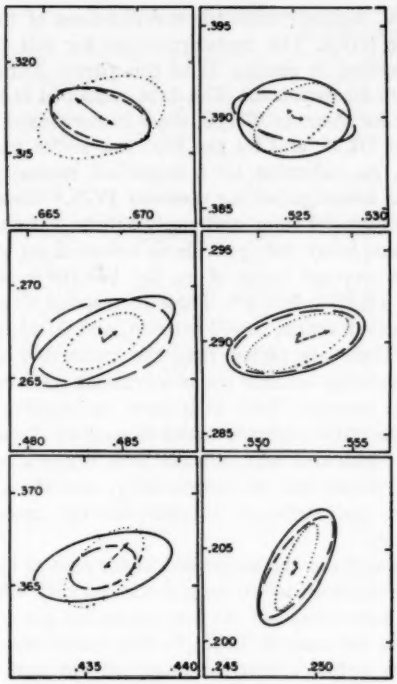

FIG. 10. Constant-luminance cross sections of discrimination ellipsoids and displacements of points of maximum luminance for colors 14 to 19.

the components of the tensor in terms of the instrument primaries.

In order to transform these results into the ICI coordinates, the transformation equations, $\{R, G, B\} = a_{ik}(X, Y, Z)$, need be known. For the first seven colors, the transformation equations were

$$R=0.28X-0.157Y-0.019Z$$
$$G=-0.36X+0.810Y-0.046Z$$
$$B=0.043X-0.061Y+0.490Z.$$

The units of Y are in foot-lamberts and those of X and Z are consistent with those units. The formulas for transforming the tensor components, c_{ik} to G_{jl}, derived from this calibration are:

$$G_{11}=0.0775c_{11}-0.1988c_{12}+0.1271c_{22}-0.0305c_{23}+0.0019c_{33}+0.0238c_{13}$$
$$G_{12}=-0.0437c_{11}+0.2820c_{12}-0.2890c_{22}+0.0564c_{23}-0.0027c_{33}-0.0239c_{13}$$
$$G_{22}=0.0244c_{11}-0.2533c_{12}+0.6565c_{22}-0.0991c_{23}+0.0038c_{33}+0.0191c_{13}$$
$$G_{23}=0.0030c_{11}-0.0080c_{12}-0.0375c_{22}+0.3990c_{23}-0.0300c_{33}-0.0755c_{13}$$
$$G_{33}=0.0003c_{11}+0.0017c_{12}+0.0022c_{22}-0.0454c_{23}+0.2390c_{33}-0.0184c_{13}$$
$$G_{13}=-0.0053c_{11}-0.0063c_{12}+0.0166c_{22}-0.1768c_{23}+0.0210c_{33}+0.1360c_{13}.$$

A slight change of primaries was made after the seventh color. The calibration equations were then found to be:

$$R=0.0780X-0.0386Y-0.0097Z$$
$$G=-0.2552X+0.5961Y-0.0401Z$$
$$B=0.0192X-0.0415Y+0.3604Z.$$

The units of Y again are foot-lamberts. These equations were used for colors 8 to 34, inclusive.

From these equations the transformation from c_{ik} to G_{jl} was made:

$$G_{11}=0.0061c_{11}-0.0398c_{12}+0.0650c_{22}-0.0098c_{23}+0.0004c_{33}+0.0030c_{13}$$
$$G_{12}=-0.0030c_{11}+0.0564c_{12}-0.1522c_{22}+0.0220c_{23}-0.0008c_{33}-0.0040c_{13}$$
$$G_{22}=0.0015c_{11}-0.0460c_{12}+0.3550c_{22}-0.0494c_{23}+0.0017c_{33}+0.0032c_{13}$$
$$G_{23}=0.0004c_{11}-0.0042c_{12}-0.0239c_{22}+0.2160c_{23}-0.0150c_{33}-0.0135c_{13}$$
$$G_{33}=0.0001c_{11}+0.0008c_{12}+0.0016c_{22}-0.0290c_{23}+0.1300c_{33}-0.0070c_{13}$$
$$G_{13}=-0.0008c_{11}-0.0007c_{12}+0.0102c_{22}-0.0929c_{23}+0.0069c_{33}+0.0279c_{13}.$$

The green primary was changed to G' for the last five colors (35 to 39). The calibration equations were redetermined with the least-squares procedure, with

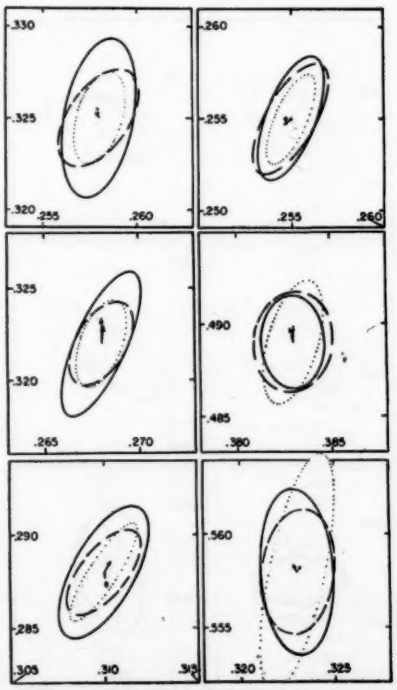

FIG. 11. Constant-luminance cross sections of discrimination ellipsoids, and displacements of points of maximum luminance for colors 20 to 25.

the result

$$R=0.07000X-0.01557Y-0.01282Z$$
$$G=-0.47550X+1.12596Y-0.06750Z$$
$$B=0.00745X-0.01828Y+0.27440Z.$$

The transformations from c_{ik} to G_{jl} for this calibration are:

$$G_{11}=0.0049c_{11}-0.0666c_{12}+0.2261c_{22}-0.0071c_{23}+0.0001c_{33}+0.0010c_{13}$$
$$G_{12}=-0.0011c_{11}+0.0862c_{12}-0.5354c_{22}+0.0171c_{23}-0.0001c_{33}-0.0014c_{13}$$
$$G_{22}=0.00024c_{11}-0.0351c_{12}+1.2678c_{22}-0.0412c_{23}+0.0003c_{33}+0.0006c_{13}$$
$$G_{23}=0.00020c_{11}-0.0134c_{12}-0.0760c_{22}+0.3102c_{23}-0.0050c_{33}-0.0040c_{13}$$
$$G_{33}=0.00016c_{11}+0.0017c_{12}+0.0046c_{22}-0.0370c_{23}+0.0753c_{33}-0.0070c_{13}$$
$$G_{13}=-0.0009c_{11}+0.0014c_{12}+0.0321c_{22}-0.1310c_{23}+0.0020c_{33}+0.0191c_{13}.$$

The values, G_{jl}, of the tensor in ICI coordinates are given in Table IV. This table also gives in Columns 3 to 5 the ICI values, X, Y, Z, of the matching field. These were computed from the observed values of R, G, B, by use of the calibration equations.

The results of the transformation to the chromaticity and $l=0.2$ log luminance coordinates are given in Table V, together with the specifications of the field in these terms. The transformations for this purpose are described in Section II of this paper. Three rows are given for each color. The data contained in the top row are for observer WRJB, those in the second are for observer DLM, and for the final row shows values of g_{11}, g_{12}, g_{22} indicated by a smoothed summary of a previous investigation for observer PGN.[4] Those data were determined from, and apply only to, chromaticity variations; hence they provide no values of g_{23}, g_{33}, and g_{31}. The average value of g_{33} for WRJB is 56×10^4, and for DLM is 64×10^4. These correspond to a value of Weber's fraction, $3\sigma_Y/Y=0.046$, and 0.043, respectively. These are higher than the value 0.02 usually cited, probably because the observations were made in a simple two-part field with dark surroundings. The variations of the separate values of g_{33} from the average are not small and may indicate that Weber's fraction is not independent of chromaticity, but the present data are not sufficient to establish the manner of dependence.

The lengths of the major and minor axes of the constant luminance ($\Delta l=0$) cross section of each ellipsoid, and the orientations of the major axes are given in the last three columns of Table V. The values shown for PGN are deduced from the smoothed summary of his results cited above. The constant-luminance cross

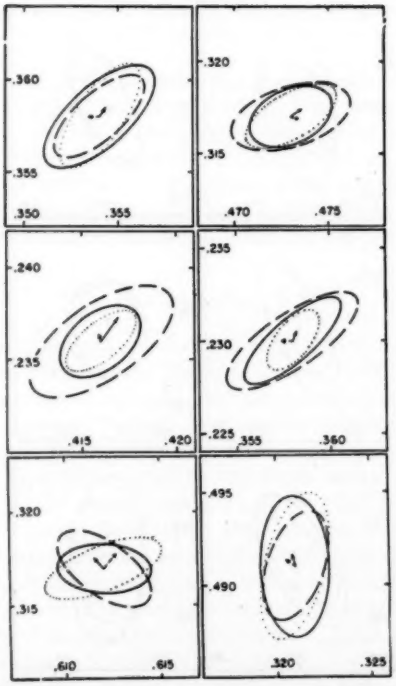

FIG. 12. Constant-luminance cross sections of discrimination ellipsoids and displacements of points of maximum luminance for colors 26 to 31.

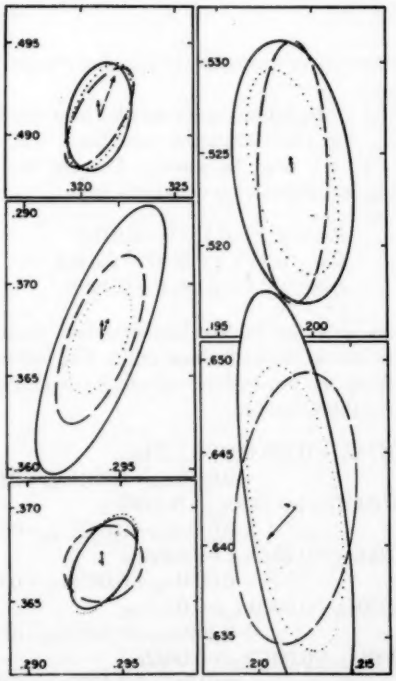

FIG. 13. Constant-luminance cross sections of discrimination ellipsoids, and displacements of points of maximum luminance for colors 32 to 36.

TABLE VI. Determinants ($\times 10^{-12}$) and minors ($\times 10^{-8}$) of discrimination ellipsoids in x, y, l and coordinates of point having maximum l.

Color	Obs.	g_0	G_{11}	G_{22}	G_{33}	G_{12}	G_{23}	Δl_m	Δx_m	Δy_m
1	WRJB	3,802	656	191	122	13	22	0.0018	0.0002	0.0003
	DLM	1,576	276	111	124	52	34	0.0028	0.0012	0.0008
2	WRJB	8,243	864	697	205	−40	33	0.0016	−0.0003	0.0002
	DLM	3,876	594	270	140	42	54	0.0019	0.0006	0.0007
3	WRJB	15,503	659	1,698	321	100	134	0.0014	0.0004	0.0006
	DLM	1,563	153	215	92	3	32	0.0024	0.0001	0.0008
4	WRJB	12,906	1,361	810	335	−65	−18	0.0016	−0.0003	−0.0001
	DLM	4,925	650	206	341	5	−36	0.0026	0.0000	−0.0003
5	WRJB	23,682	1,166	983	509	−38	55	0.0015	−0.0001	0.0002
	DLM	7,930	575	420	276	10	70	0.0019	0.0007	0.0005
6	WRJB	30,733	812	4,495	496	7	38	0.0013	0.0000	0.0001
	DLM	21,204	1,260	2,394	594	−444	−684	0.0017	−0.0012	−0.0019
7	WRJB	31,160	1,168	1,355	747	4	215	0.0016	0.0000	0.0004
	DLM	56,912	3,205	1,829	629	−155	355	0.0011	−0.0003	0.0006
8	WRJB	12,222	584	1,305	230	−110	138	0.0014	−0.0007	0.0008
9	WRJB	67,772	2,158	986	3,590	−704	700	0.0023	−0.0004	0.0004
	DLM	13,352	816	271	1,440	8	128	0.0033	0.0000	0.0003
11	WRJB	36,616	1,986	554	1,462	−290	338	0.0020	−0.0004	0.0005
	DLM	19,189	1,121	297	1,163	193	3	0.0025	0.0004	0.0000
12	WRJB	18,569	2,087	1,488	404	155	155	0.0015	0.0006	0.0006
	DLM	44,843	3,377	1,630	971	−430	115	0.0015	−0.0006	0.0002
13	WRJB	197,579	3,955	4,108	3,874	−329	−553	0.0014	−0.0001	−0.0002
	DLM	546,989	9,969	10,256	9,239	−3,171	−3,070	0.0013	0.0004	−0.0004
14	WRJB	12,332	1,061	444	592	−350	256	0.0022	−0.0013	0.0010
	DLM	15,377	1,218	365	773	255	−144	0.0022	0.0007	−0.0004
15	WRJB	5,751	855	359	251	87	62	0.0021	0.0007	0.0005
	DLM	3,358	534	178	145	−14	48	0.0021	−0.0002	0.0007
16	WRJB	205,255	6,914	5,024	2,009	586	189	0.0010	0.0003	0.0001
	DLM	140,551	6,813	4,367	2,895	2,686	2,676	0.0014	0.0013	0.0013
17	WRJB	8,503	791	387	242	66	−11	0.0017	0.0005	−0.0001
	DLM	16,791	1,765	451	442	−88	121	0.0016	−0.0003	0.0004
18	WRJB	6,808	867	256	299	23	92	0.0021	0.0002	0.0006
	DLM	13,208	1,269	476	461	419	146	0.0018	0.0018	0.0006
19	WRJB	11,773	575	1,372	374	−56	−83	0.0018	−0.0003	−0.0004
	DLM	29,271	1,223	2,246	767	−37	−157	0.0016	−0.0001	0.0003
20	WRJB	7,554	284	1,463	161	−20	37	0.0015	−0.0002	0.0003
	DLM	13,185	658	1,053	382	−29	141	0.0017	−0.0001	0.0006
21	WRJB	13,988	472	2,196	326	−2	294	0.0015	0.0000	0.0014
	DLM	48,840	1,458	3,002	908	72	656	0.0014	0.0001	0.0010
22	WRJB	15,646	1,020	2,039	238	−154	16	0.0012	−0.0008	0.0001
	DLM	58,052	3,096	3,863	764	208	406	0.0012	0.0003	0.0006
23	WRJB	26,847	806	2,909	531	−159	−120	0.0014	−0.0004	−0.0003
	DLM	18,956	789	1,692	549	−127	−8	0.0017	−0.0004	0.0000
24	WRJB	28,510	815	2,134	525	−15	350	0.0014	0.0000	0.0009
	DLM	14,586	677	1,238	282	−36	168	0.0014	−0.0002	0.0008
25	WRJB	5,321	226	1,095	119	−17	51	0.0015	−0.0002	0.0006
	DLM	21,792	936	2,484	212	48	−60	0.0010	0.0002	0.0003

TABLE VI.—*Continued.*

Color	Obs.	g_0	G_{11}	G_{22}	G_{33}	G_{12}	G_{23}	Δl_m	Δx_m	Δy_m
26	WRJB	27,174	2,009	2,253	287	77	147	0.0010	0.0003	0.0005
	DLM	98,218	5,304	4,684	839	−570	−34	0.0009	−0.0006	0.0000
27	WRJB	7,195	807	695	277	122	185	0.0020	0.0009	0.0013
	DLM	3,413	507	327	128	−7	30	0.0019	−0.0001	0.0004
28	WRJB	33,029	2,176	672	935	−293	316	0.0017	−0.0005	0.0006
	DLM	27,041	1,751	1,271	653	316	298	0.0016	0.0007	0.0008
29	WRJB	57,532	3,068	1,700	931	570	380	0.0016	0.0006	0.0004
	DLM	32,144	3,234	1,128	344	140	−48	0.0010	0.0004	−0.0001
30	WRJB	40,160	2,629	2,324	724	126	276	0.0013	0.0002	0.0005
	DLM	22,039	2,876	1,580	293	−165	−11	0.0012	−0.0006	0.0000
31	WRJB	24,272	794	3,602	216	−40	160	0.0009	−0.0002	0.0007
	DLM	71,839	2,267	6,521	481	−291	218	0.0008	−0.0005	0.0004
32	WRJB	48,732	1,320	4,799	444	−36	426	0.0010	−0.0001	0.0009
	DLM	74,286	2,925	9,569	746	570	1,660	0.0010	0.0008	0.0022
33	WRJB	2,231	279	1,208	32	2	27	0.0012	0.0001	0.0010
	DLM	10,948	648	2,486	140	56	140	0.0011	0.0004	0.0011
34	WRJB	62,192	1,856	4,287	807	−72	475	0.0011	−0.0001	0.0007
	DLM	64,830	2,926	2,701	559	−122	137	0.0009	−0.0002	0.0002
35	WRJB	750	79	400	19	−1	10	0.0016	−0.0001	0.0008
	DLM	3,212	115	1,620	56	−4	28	0.0013	−0.0001	0.0007
36	WRJB	805	91	1,121	11	−14	−16	0.0012	−0.0015	−0.0017
	DLM	1,261	138	721	19	−10	33	0.0012	−0.0006	0.0002
37	WRJB	15,850	850	4,342	157	−150	−12	0.0010	−0.0009	−0.0001
	DLM	2,727	211	892	52	−8	−5	0.0014	−0.0002	−0.0001
38	WRJB	97,352	594	13,690	1,478	42	−1,994	0.0012	0.0000	−0.0017
	DLM	44,574	191	5,438	2,126	−70	−976	0.0022	−0.0001	−0.0010
39	WRJB	15,185	332	3,983	183	4	3	0.0011	0.0000	0.0000
	DLM	16,808	544	1,616	345	−140	−68	0.0014	−0.0006	−0.0003

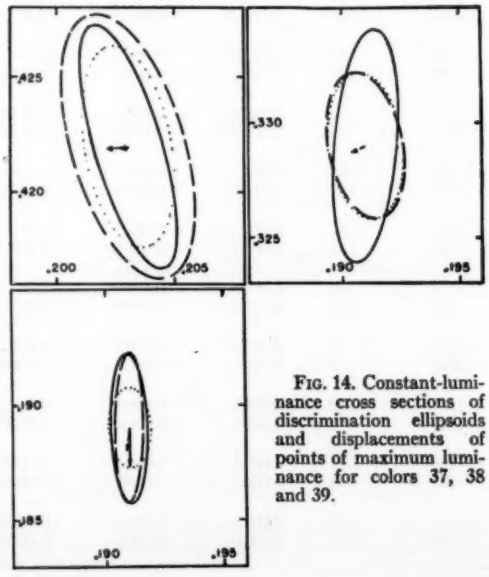

FIG. 14. Constant-luminance cross sections of discrimination ellipsoids and displacements of points of maximum luminance for colors 37, 38 and 39.

sections are shown for the three observers in Figs. 8 to 14. Figure 15 shows all of these for WRJB, each enlarged ten times about its center, on the ICI chromaticity diagram.

Values of the determinant, g_0, and its minors, G_{11}, G_{22}, G_{33}, G_{13}, G_{23}, sufficient to determine the axes of the ellipsoids and the displacement of the point of maximum luminance on each ellipsoid are given in the first six columns of Table VI. The displacement of the point of maximum luminance on each ellipsoid, relative to its center, is given in the last three columns of Table VI. The chromaticity displacement of this point, as well as the central constant-luminance cross section of the ellipsoids, are shown in Figs. 8 to 14. In general, the displacements, Δx_m and Δy_m are small compared to the dimensions of the ellipsoid of discrimination. This indicates that the asymmetries of those ellipsoids with respect to luminance are relatively insignificant. The most notable exceptions are in the red region, colors 9, 11, and 14.

Table VII gives the length and orientation of the principal axes of the ellipsoids in the frame of reference,

TABLE VII. Lengths and orientations of principal axes of discrimination ellipsoids in x, y, $\frac{1}{3}\log Y$ coordinates.

Color	x	y	Y	a_1	θ_1	ϕ_1	a_2	θ_2	ϕ_2	a_3	θ_3	ϕ_3
1	0.580	0.341	5.4	0.00418	91°	−83°	0.00224	71°	5°	0.00173	160°	19°
				0.00441	103°	−75°	0.00310	40°	−1°	0.00184	127°	25°
2	0.422	0.330	5.8	0.00396	90°	50°	0.00190	62°	−39°	0.00146	151°	−39°
				0.00426	84°	63°	0.00218	60°	−30°	0.00173	149°	−15°
3	0.364	0.505	4.8	0.00356	84°	24°	0.00164	108°	−64°	0.00138	20°	−83°
				0.00411	81°	32°	0.00272	63°	−63°	0.00225	151°	−41°
4	0.370	0.218	2.75	0.00385	84°	55°	0.00164	19°	−26°	0.00141	107°	−36°
				0.00392	91°	66°	0.00265	172°	−39°	0.00137	82°	−24°
5	0.492	0.446	3.35	0.00232	90°	58°	0.00193	80°	−34°	0.00145	171°	−26°
				0.00275	84°	69°	0.00231	68°	−23°	0.00176	157°	−9°
6	0.308	0.424	2.95	0.00400	90°	18°	0.00127	4°	−78°	0.00112	93°	−72°
				0.00408	97°	30°	0.00132	26°	89°	0.00127	69°	−50°
7	0.446	0.450	5.8	0.00238	80°	38°	0.00169	54°	−60°	0.00141	142°	−39°
				0.00238	86°	−89°	0.00184	75°	2°	0.00096	164°	−12°
8	0.319	0.581	4.1	0.00330	82°	5°	0.00222	104°	82°	0.00124	165°	67°
9	0.682	0.302	6.1	0.00248	31°	−55°	0.00172	120°	−70°	0.00090	97°	24°
				0.00330	7°	−19°	0.00266	95°	−67°	0.00098	95°	23°
11	0.650	0.292	6.2	0.00241	64°	−82°	0.00197	34°	57°	0.00110	109°	0°
				0.00265	137°	−86°	0.00223	49°	−83°	0.00122	92°	6°
12	0.314	0.183	0.87	0.00423	86°	50°	0.00144	171°	75°	0.00120	100°	−38°
				0.00314	95°	59°	0.00159	31°	−22°	0.00094	120°	−34°
13	0.221	0.148	0.55	0.00183	103°	44°	0.00138	173°	60°	0.00089	86°	−45°
				0.00191	112°	45°	0.00117	22°	42°	0.00061	91°	−46°
14	0.668	0.317	2.47	0.00338	63°	−64°	0.00188	44°	58°	0.00141	121°	8°
				0.00307	113°	−70°	0.00207	24°	−89°	0.00127	83°	17°
15	0.484	0.267	1.07	0.00437	83°	60°	0.00203	170°	89°	0.00148	96°	−28°
				0.00406	90°	77°	0.00244	49°	−13°	0.00173	139°	−11°
16	0.436	0.366	3.3	0.00218	86°	52°	0.00107	120°	−35°	0.00095	30°	−45°
				0.00281	66°	53°	0.00121	62°	−52°	0.00078	141°	−3°
17	0.525	0.390	2.88	0.00306	97°	−89°	0.00213	95°	−2°	0.00166	8°	−59°
				0.00332	85°	−76°	0.00173	37°	20°	0.00134	126°	10°
18	0.552	0.290	4.6	0.00371	86°	72°	0.00223	29°	−22°	0.00146	118°	−16°
				0.00337	67°	76°	0.00183	77°	−20°	0.00141	148°	43°
19	0.249	0.204	0.92	0.00384	94°	29°	0.00176	18°	4°	0.00136	93°	−61°
				0.00326	86°	34°	0.00162	−9°	−10°	0.00111	82°	−56°
20	0.258	0.325	1.95	0.00446	89°	7°	0.00182	73°	−80°	0.00142	183°	−75°
				0.00320	84°	32°	0.00189	46°	−64°	0.00143	136°	−51°
21	0.268	0.322	2.02	0.00414	82°	16°	0.00165	49°	−81°	0.00123	133°	−65°
				0.00275	77°	26°	0.00148	54°	−73°	0.00111	142°	−48°
22	0.310	0.288	1.85	0.00406	92°	31°	0.00188	66°	−59°	0.00105	156°	−63°
				0.00327	85°	41°	0.00121	49°	−53°	0.00104	139°	−44°
23	0.255	0.255	1.32	0.00350	94°	21°	0.00149	35°	−64°	0.00117	125°	−71°
				0.00337	93°	30°	0.00176	22°	−53°	0.00122	112°	−61°
24	0.383	0.489	7.3	0.00278	78°	0°	0.00169	92°	89°	0.00126	168°	−18°
				0.00295	81°	7°	0.00214	98°	96°	0.00130	168°	−33°
25	0.323	0.558	6.20	0.00454	87°	−4°	0.00205	97°	86°	0.00147	9°	108°
				0.00340	91°	7°	0.00205	86°	97°	0.00097	5°	−67°
26	0.354	0.358	6.31	0.00359	87°	42°	0.00167	85°	−48°	0.00101	174°	−24°
				0.00293	93°	48°	0.00132	68°	−41°	0.00083	158°	−51°

TABLE VII.—*Continued.*

Color	x	y	Y	a_1	θ_1	ϕ_1	a_2	θ_2	ϕ_2	a_3	θ_3	ϕ_3
27	0.416	0.236	4.19	0.00436	79°	47°	0.00192	32°	−61°	0.00140	120°	−36°
				0.00452	89°	54°	0.00217	50°	−37°	0.00174	140°	−34°
28	0.612	0.317	6.18	0.00263	82°	−80°	0.00177	38°	30°	0.00118	124°	−6°
				0.00287	92°	−55°	0.00206	50°	33°	0.00102	139°	37°
29	0.473	0.317	10.74	0.00257	77°	61°	0.00142	81°	−31°	0.00115	157°	27°
				0.00328	88°	73°	0.00168	99°	−17°	0.00101	9°	−31°
30	0.358	0.230	3.66	0.00332	86°	47°	0.00138	28°	−51°	0.00109	118°	−39°
				0.00427	92°	56°	0.00148	67°	−33°	0.00107	156°	−39°
31	0.321	0.491	4.91	0.00384	87°	−3°	0.00180	93°	87°	0.00092	176°	−52°
				0.00312	89°	18°	0.00161	76°	−73°	0.00074	166°	−68°
32	0.294	0.367	5.71	0.00318	85°	10°	0.00159	82°	−81°	0.00090	170°	−46°
				0.00397	81°	25°	0.00123	75°	−68°	0.00075	162°	−36°
33	0.294	0.367	0.31	0.00782	89°	21°	0.00230	85°	−69°	0.00117	176°	−58°
				0.00507	86°	21°	0.00172	95°	−69°	0.00110	177°	29°
34	0.385	0.400	6.96	0.00286	85°	26°	0.00141	60°	−67°	0.00099	149°	−55°
				0.00446	90°	52°	0.00192	84°	−38°	0.00091	172°	−57°
35	0.199	0.524	1.45	0.00735	88°	−7°	0.00314	90°	83°	0.00158	2°	23°
				0.00709	91°	0°	0.00190	94°	90°	0.00131	5°	102°
36	0.212	0.642	1.38	0.01188	92°	−7°	0.00315	103°	83°	0.00094	13°	81°
				0.00758	85°	5°	0.00326	84°	−86°	0.00113	6°	130°
37	0.203	0.419	1.54	0.00544	90°	−17°	0.00184	111°	73°	0.00079	21°	74°
				0.00591	90°	−16°	0.00237	95°	74°	0.00137	5°	73°
38	0.171	0.189	2.55	0.00379	99°	−2°	0.00109	9°	−19°	0.00077	88°	88°
				0.00358	105°	1°	0.00204	15°	−5°	0.00065	88°	91°
39	0.191	0.329	1.67	0.00513	90°	4°	0.00144	92°	−86°	0.00110	2°	−86°
				0.00332	92°	−9°	0.00188	121°	82°	0.00123	31°	78°

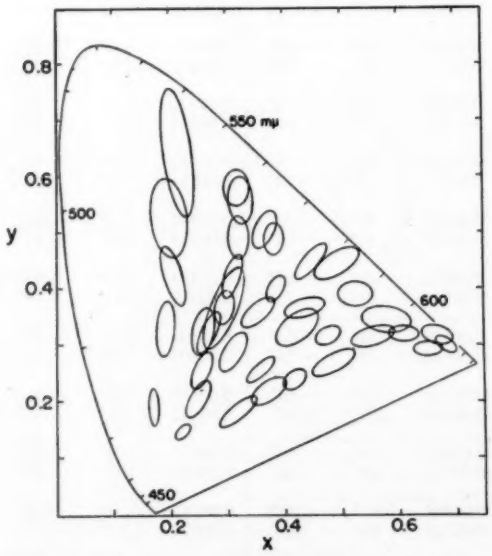

FIG. 15. Cross sections at constant luminance of discrimination ellipsoids for WRJB, enlarged ten times on chromaticity diagram.

x, y, and l, where the last is the common logarithm of luminance plotted to a scale only one fifth as great as the scale of the chromaticity coordinates. The values of colatitude, longitude, and length are given for each of the axes in turn. The colatitude is measured from the log$Y$ axis and the longitude is measured from the (y, log$Y$) plane. The values shown specify the shapes and orientations of all the standard deviation ellipsoids. From these data the models portrayed in the color plate were prepared.

The influence of luminance on the results is indicated by colors 32 and 33, which differ only in luminance. The principal effect of decrease of luminance appears to be a proportionate increase in the chromaticity components of all radii of the discrimination ellipsoid, accompanied by a much smaller percentage increase of the luminance components.

The influence of luminance was studied in more detail for three other chromaticities, blue (19a), green (31a), and white (34a). In this series of experiments, the luminance was varied by placing non-selective gray filters in the eyepiece. Since these absorbed light equally from both halves of the colorimeter field, the average

TABLE VIII. Influence of luminance variation on precision of color-matching.

Color	Obs.	R	G	B	σ_r	σ_g	σ_b	ρ_{gb}	ρ_{br}	ρ_{rg}
No. 19*a*	WRJB	0.0289	0.154	0.904	0.000580	0.00443	0.02075	−0.59	0.56	−0.35
	DLM	0.0303	0.152	0.931	0.000452	0.00262	0.01568	−0.38	0.42	−0.18
+0.30	WRJB	0.0145	0.0773	0.454	0.0002635	0.002855	0.00844	−0.47	0.33	−0.47
	DLM	0.0152	0.0768	0.467	0.000296	0.001822	0.00925	−0.18	0.11	−0.18
+1.00	WRJB	0.00289	0.0154	0.0904	0.0001103	0.000848	0.003018	−0.53	0.60	−0.41
	DLM	0.00303	0.0153	0.0931	0.0000819	0.000606	0.003465	−0.41	0.46	−0.39
+1.55	WRJB	0.000816	0.00435	0.0255	0.0000439	0.000572	0.001178	−0.53	0.47	−0.47
	DLM	0.000856	0.00432	0.0263	0.0000431	0.000362	0.001880	−0.27	0.45	−0.39
No. 31*a*	WRJB	0.0392	1.94	0.814	0.001607	0.01782	0.01708	−0.22	0.27	−0.11
	DLM	0.0383	1.96	0.810	0.001176	0.01864	0.01747	0.12	0.15	−0.18
+0.30	WRJB	0.0197	0.975	0.410	0.000801	0.01140	0.00867	−0.08	0.38	−0.10
	DLM	0.0192	0.984	0.408	0.000712	0.01019	0.00887	0.04	−0.13	−0.06
+1.00	WRJB	0.00392	0.194	0.0814	0.0002100	0.002645	0.00490	−0.03	0.43	−0.45
	DLM	0.00383	0.196	0.0810	0.0001713	0.002355	0.00483	−0.23	0.42	−0.37
+1.55	WRJB	0.001106	0.0548	0.02297	0.0000753	0.000808	0.002487	−0.32	0.05	−0.24
	DLM	0.001080	0.0553	0.02285	0.0000842	0.000960	0.002772	−0.42	0.37	−0.35
No. 34*a*	WRJB	0.215	2.31	1.208	0.002975	0.02105	0.02385	0.23	0.56	0.35
	DLM	0.227	2.30	1.200	0.00304	0.0195	0.01946	0.23	0.21	0.09
+0.30	WRJB	0.1079	1.160	0.604	0.00128	0.01197	0.01702	−0.13	0.32	0.29
	DLM	0.114	1.154	0.606	0.001706	0.01367	0.01479	0.10	0.06	0.46
+0.78	WRJB	0.03575	0.384	0.205	0.000498	0.00432	0.00766	−0.06	0.30	0.00
	DLM	0.03775	0.382	0.1993	0.000612	0.00475	0.00538	0.11	0.27	0.15
+1.10	WRJB	0.01710	0.1835	0.0960	0.000323	0.00205	0.00419	0.02	0.53	−0.05
	DLM	0.01806	0.1828	0.0954	0.000264	0.00260	0.00290	0.03	0.25	−0.09
+1.55	WRJB	0.00606	0.0651	0.03405	0.0001219	0.001386	0.003285	−0.36	0.31	0.03
	DLM	0.00640	0.0649	0.03385	0.0001286	0.000983	0.002222	−0.26	0.14	0.16
+2.10	WRJB	0.001710	0.01835	0.00960	0.0000763	0.000625	0.001105	−0.52	0.60	−0.36
	DLM	0.001806	0.01828	0.00954	0.0000668	0.000499	0.000961	0.00	0.28	−0.02

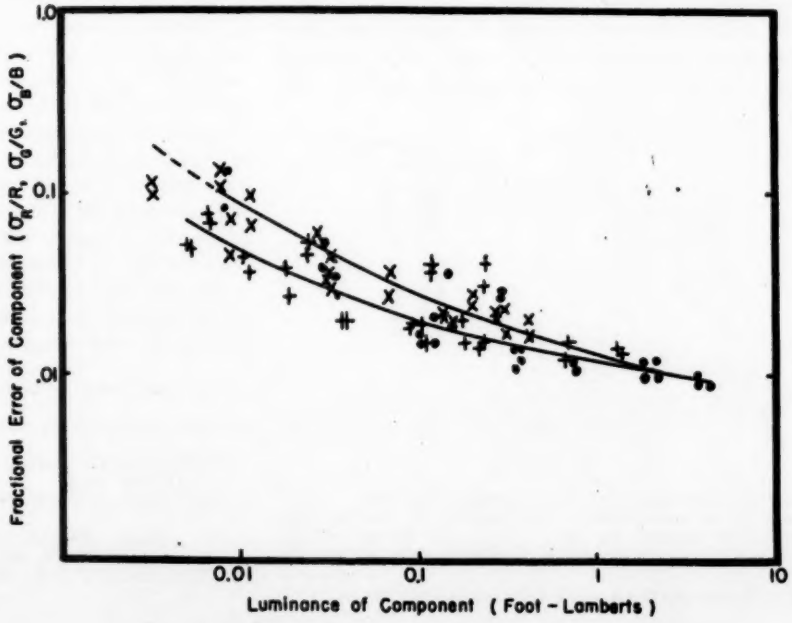

FIG. 16. Fractional errors of colorimeter primaries as functions of luminance contributed by each primary to match:

+ = Values for red primary
●●● = Values for green primary
××× = Values for blue primary

Lower curve fits values for red primary.

Upper curve fits values for green and blue primaries.

TABLE IX. Luminance of components and fractional error of color-matching

Color	Obs.	Y_r	r/R	Y_g	g/G	Y_b	b/B
No. 19a	WRJB	0.1814	0.020	0.302	0.029	0.312	0.023
	DLM	0.1887	0.015	0.299	0.017	0.321	0.017
+0.30	WRJB	0.0910	0.018	0.152	0.037	0.157	0.019
	DLM	0.0952	0.019	0.151	0.024	0.161	0.020
+1.00	WRJB	0.0181	0.038	0.0302	0.055	0.031	0.033
	DLM	0.0190	0.027	0.0299	0.039	0.032	0.037
+1.55	WRJB	0.00512	0.052	0.0086	0.132	0.0088	0.046
	DLM	0.00536	0.050	0.0085	0.084	0.0091	0.072
No. 31a	WRJB	0.245	0.041	3.817	0.009	0.281	0.021
	DLM	0.240	0.031	3.853	0.010	0.279	0.022
+0.30	WRJB	0.123	0.041	1.917	0.012	0.141	0.021
	DLM	0.120	0.037	1.936	0.010	0.141	0.022
+1.00	WRJB	0.0246	0.054	0.382	0.014	0.028	0.060
	DLM	0.0240	0.045	0.385	0.012	0.028	0.060
+1.55	WRJB	0.0069	0.068	0.108	0.015	0.0079	0.108
	DLM	0.0068	0.078	0.109	0.017	0.0079	0.134
No. 34a	WRJB	1.347	0.014	4.54	0.009	0.417	0.020
	DLM	1.422	0.013	4.52	0.009	0.414	0.016
+0.30	WRJB	0.677	0.012	2.28	0.010	0.208	0.028
	DLM	0.715	0.015	2.27	0.012	0.209	0.024
+0.78	WRJB	0.224	0.014	0.755	0.011	0.0708	0.037
	DLM	0.237	0.016	0.750	0.012	0.0688	0.027
+1.10	WRJB	0.1072	0.019	0.361	0.011	0.0332	0.044
	DLM	0.1131	0.015	0.359	0.014	0.0329	0.030
+1.55	WRJB	0.0380	0.020	0.128	0.021	0.0118	0.097
	DLM	0.0401	0.020	0.128	0.015	0.0117	0.066
+2.10	WRJB	0.01072	0.045	0.036	0.034	0.0033	0.115
	DLM	0.01131	0.037	0.036	0.027	0.0033	0.099

settings for match did not change significantly with luminance. The standard deviations of the settings are therefore directly indicative of the effect of luminance on the precision of color-matching. The results in terms of the amounts of the instrument primaries are given in Table VIII. In order to determine the major effects for sufficient number of luminances, only sixty matches were made by each observer at each luminance. These are sufficient for quite good evaluations of standard deviations of the instrument primaries but are not sufficient for reliable correlation coefficients. Although some indications were obtained that the correlation coefficients change with luminance, the results are too indefinite for publication. The influence of luminance will be exhibited only in terms of the standard deviations of the instrument readings.

Several ways of plotting the effects of luminance variation have been tried. The most interesting is shown in Fig. 16, in which the ratio between the standard deviation and the amount of each primary is plotted against the luminance contribution of that primary to the three-color mixture. The data from which Fig. 16 was plotted are given in Table IX. The points shown in Fig. 16 for the red primary appear to be consistently lower than those for the green and blue primaries. The best curves drawn through these three sets of points are suggestively similar to the curves of $\Delta I/I$ found by Hecht, Peskin, and Patt for spectral radiation.

Refractive Indices of Water Vapor and Carbon Dioxide at Low Pressure

KENNETH B. NEWBOUND*
George Eastman Research Laboratories of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts†
(Received May 20, 1949)

The Fabry-Perot interferometer has been used to obtain the dispersion curves for water vapor and for carbon dioxide in the wave-length range 2500A to 8600A at pressures below 20 mm of mercury and at a temperature of 20°C. The results have been formulated to permit computation of the refractive index of air containing these substances. The presence of water vapor in air used for precision wave-length measurements is undesirable, due to possible formation of an adsorbed film of water on the etalon mirrors. This has been observed for evaporated aluminium mirrors and for evaporated silver mirrors.

Temperature variation causes the effective thickness of the common form of Fabry-Perot etalon to vary in much greater degree than is predicted by the expansion coefficient of the spacer material. The explanation of this effect leads to the conclusion that the etalon pressure coefficients will show similar anomalies. It is possible to explain qualitatively the discrepancies between the measurements of the refractive index of air made by Meggers and Peters and those by more recent observers.

IN general, interferometric determinations of wave-length in terms of the primary standard require corrections in order to reduce them to the values for the standard conditions of dry air at 15°C and 760 mm containing 0.03 percent by volume of carbon dioxide. The corrections are readily made provided dry, carbon dioxide free air is used. However, it is not always convenient to use air freed from these substances, and the effects of their presence may have to be considered if high precision is required. The work here described provides corrections for the presence of known amounts of water vapor, carbon dioxide, or both.

A knowledge of the refractive index of air is essential in order to make use of observed wave-lengths for the construction of energy level diagrams. Refractive index data of various observers have been in some disagreement. The situation has been discussed by Sawyer[1] who does not point out, however, that the data of Barrell and Sears[2] and of Pérard[3] were limited to the visible region of the spectrum, while the experimental details of the work of Kösters and Lampe[4] have not been published. The dispersion equations of Barrell and Sears and of Pérard are in good agreement in the visible region for which data were obtained, while the data of Meggers and Peters[5] exhibit less dispersion and generally lower indices. In the case of the latter workers, no correction was made for the change in dimensions of their Fabry-Perot interferometers when evacuated and when filled with air at pressures near 760 mm of mercury. Elementary considerations of the bulk modulus of Invar would seem to suggest that this effect is too small to explain the discrepancies. The interferometer mountings used for the present work were of similar construction to those used by Meggers and Peters and some observations were made, which, it is believed explain the discrepancies.

On the basis of electromagnetic theory, the refractive index of a mixture of gases is readily computed, provided the indices of the constituents are known. The refractive index of water vapor has been measured in the visible by the Cuthbertsons,[6] working at pressures of the order of an atmosphere and temperatures exceeding 100°C. Barrell and Sears[2] have measured the index of moist air in the visible. They measured, in effect, the slight change in the refractive index of air when a small amount of water vapor was added, and hence, their results do not have high precision. In order to duplicate as closely as possible the desired conditions and still attain reasonable accuracy of measurement, the work here described on water vapor used pressures of less than 15 mm and temperatures near 20°C. Further, the range of the data was that in which

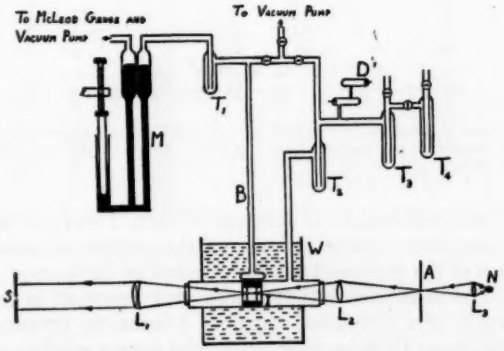

FIG. 1. Schematic diagram of apparatus. *A*, aperture. *B*, "thermometer" arm. *D*, discharge tube. *I*, interferometer. L_1–L_3, lenses. *S*, spectrograph slit. T_1–T_4, traps. *M*, mercury manometer. *W*, constant temperature bath. *N*, light source.

* Now at the University of Alberta, Edmonton, Alberta, Canada.

† Work carried out in the Spectroscopy Laboratory under the direction of George R. Harrison.

[1] R. A. Sawyer, *Experimental Spectroscopy* (Prentice Hall, Inc., New York, 1944), p. 210.

[2] H. Barrell and J. E. Sears, Phil. Trans. Royal Soc. **238A**, 1 (1939).

[3] A. Pérard, Trav. Bur. Int. Poids Mes. **19**, 1 (1934).

[4] Kösters and Lampe, Physik. Zeits. **35**, 223 (1934).

[5] W. F. Meggers and C. G. Peters, Bull. Bur. Stand. **327**, 697 (1918).

[6] C. and M. Cuthbertson Phil. Trans. Royal Soc. **213A** 1 (1913).

precision spectroscopic measurements are customarily made.

Koch[7] has made measurements on the refractive index of carbon dioxide covering the spectral region 2379A to 6709A and using pressures near atmospheric. The data of other observers[8,9] were limited to the visible. The work here described was done at pressures of less than 20 mm and at a temperature near 20°C, covering the same wave-length region as for the water vapor.

EXPERIMENTAL

Fabry-Perot etalons consisting of aluminized crystal-quartz mirrors separated by fused quartz spacers were employed for all measurements except a few, where silvered mirrors were used. The mirrors were 6 cm in diameter and about 1 cm thick, the coated and uncoated surfaces making a slight angle with each other so that secondary fringe systems could be thrown to one side of the main system. Three optically flat bosses symmetrically located on each end of the spacers made optical contact with each mirror. The mirror-spacer assembly was held in a brass cylinder by means of three springs at each end of the cylinder. The mirror surfaces were adjusted to parallelism by variation in the flexure of one or more of these springs. Two spacers were used,

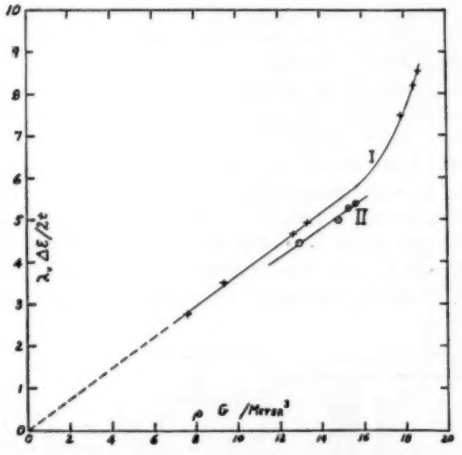

FIG. 2. Fractional change in optical thickness of etalon $\lambda_r\Delta\epsilon/2t$ *vs.* water vapor density ρ at 22.25°C; *I*: $t=14.983$ mm; *II*: $t=24.940$ mm.

of nominal lengths 15 mm and 25 mm. Theory of the Fabry-Perot interferometer and the method of reduction of the patterns have been treated by Meissner.[10]

The assembled interferometer was mounted in the center of a cylindrical chamber 3 inches in diameter and about 15 inches long, fitted with quartz windows at each end and surrounded for 12 inches of its length by a water bath whose temperature was thermostatically controlled, such that the periodic fluctuations were less than 0.05°C. The chamber (see Fig. 1) was connected to one arm of the mercury manometer *M* and a vacuum pump. The traps and stopcocks indicated facilitated the filling and control of the gaseous content of the chamber.

The optical train, also indicated in Fig. 1, consisted of the light source, *N* (Geissler tube containing neon, or mercury vapor lamp) which was focused by the quartz lens L_3 on a circular aperture *A*, 3 mm in diameter. The focusing of *A* on the etalon *I* by means of the quartz lens L_2 limited the effective area of the interferometer so that the effects of any slight irregularities of the interferometer plates were reduced. The fringes produced by the interferometer were focused on the slit *S* of a stigmatic spectrograph employing a 35-foot concave grating (15,000 lines per inch) by the lens L_1, a quartz-fluorite achromat of about 50 cm focal length which was so placed that the interferometer was imaged in the vicinity of the grating. A wave-length range of slightly over 1500A (first order) could be photographed on a 20-inch plate.

The vapor pressures in the chamber were measured by means of the manometer *M*. The arms were filled with redistilled mercury, and provision was made for raising the levels slightly before taking a reading so that rising columns of mercury were being used. The portions of the Pyrex tubing where readings were taken were 20 mm in diameter, minimizing the effects of any slight irregularities in the bore. The reading of the top of the mercury meniscus was first obtained by means of a shadow projection on a microscope scale (0.05-mm divisions). The viewing assembly was then rotated so that a brass scale appeared in the field of view. The differences in mercury levels obtained were accurate to about 0.002 mm. All pressure determinations were corrected for temperature and gravity.

The temperature of the chamber was observed by means of a Beckmann thermometer suspended in the water bath near the center of the chamber. The calibration of the Beckmann was determined by comparison with the pressure of saturated water vapor in the arm *B*, the temperature of the water bath being between one and three degrees less than room temperature. This calibration was checked during the course of the experiments and did not vary by more than the experimental errors of measurement, 0.003°C.

THE REFRACTIVE INDEX OF WATER VAPOR

The principal data used for the computation of the refractive index of water vapor were obtained, for each wave-length region, from four plates, two for each etalon spacer. Each plate carried four exposures, separated by sideways motion of the plate between exposures. First a vacuum exposure was made, followed by two exposures with the chamber filled with water

[7] P. P. Koch, Nova Acta Reg. Upsaliensis 2, No. 5 (1909).
[8] E. Stoll, Ann. d. Physik 69, 81 (1922).
[9] C. and M. Cuthbertson, Proc. Roy. Soc. London A97, 152 (1920).
[10] K. W. Meissner, J. Opt. Soc. Am. 31, 405 (1941).

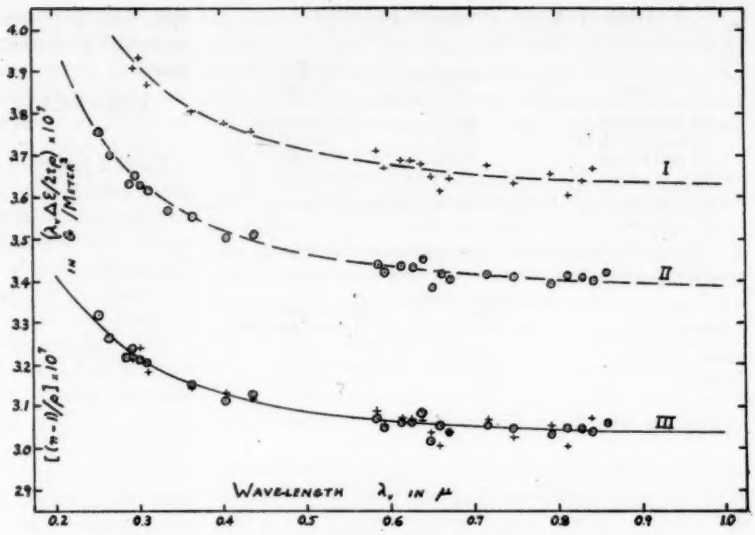

FIG. 3. *I*: Fractional change in optical thickness of etalon, $\lambda_v\Delta\epsilon/2l\rho$ *vs.* wave-length, $t=14.983$ mm. *II*: Fractional change in optical thickness of etalon $\lambda_v\Delta\epsilon/2l\rho$ *vs.* wave-length, $t=24.940$ mm. *III*: Specific refractive index $(n-1)/\rho$ of water vapor at 22.25°C *vs.* wave-length.

TABLE I. Specific refractive indices of water vapor and of water.

Wave-length in microns	Specific refractive indices in (g/meter³)×10⁷ Equation (6)	Cuthbertsons	Barrell & Sears	water
0.2	3.512±0.032	—	—	3.793
0.4	3.127±0.0025	3.201	3.21	3.173
0.6	3.067±0.0028	3.107	3.12	3.083
0.8	3.046±0.0032	(3.074)	(3.09)	3.053
1.0	3.036±0.0035	(3.059)	(3.07)	3.039
∞	(3.020±0.0041)	(3.032)	(3.04)	(3.016)

vapor at pressures between 6 and 15 mm of mercury. Finally a second vacuum exposure was made. Additional data were obtained using the 15-mm etalon for vapor pressures nearer saturation. The source of water vapor was a quantity of distilled water contained in the trap T_4. When the vapor pressure in the chamber had reached the desired value, T_4 was isolated and the system allowed to stand for half an hour before the exposure was begun. During the half-hour exposure, the pressure and temperature were read at five-minute intervals. The standard deviations of these sets of readings were less than 0.008 mm of mercury and 0.006°C. The mean water vapor density was computed from a table[11] of saturated water vapor *versus* pressure and temperature.

The determination of the refractive index of water vapor was found to be complicated by a film of adsorbed water on the mirror surfaces of the etalon. Correction for this film was simplified when the change in optical thickness was found to be directly proportional to the water vapor density for vapor pressures below about 70 percent of the saturation pressure. This feature is illustrated in Fig. 2 where the fractional change in optical thickness of the etalon is plotted against water vapor density for the two etalon thicknesses 14.983 mm and 24.940 mm. Each point plotted represents the mean value for six wave-lengths in the range 5854A to 6508A.

Making use of the fundamental equation of interferometry and the relation between wave-length and refractive index, we may write for the linear portion of these curves

$$\lambda_v\Delta\epsilon=2l(n-1)+2K\rho \qquad (1)$$

TABLE II. Factors for computation of the refractive index of moist air.

Temperature θ°C	Wave-length λ_v microns 0.2	0.3	0.4	$F(\theta)\times10^7$ 0.5	0.6	0.8	1.0
10	7.37	5.82	5.56	5.49	5.47	5.45	5.45
12	8.39	6.63	6.33	6.25	6.22	6.21	6.20
14	9.52	7.51	7.18	7.09	7.06	7.04	7.03
16	10.77	8.51	8.13	8.03	7.99	7.97	7.96
18	12.17	9.61	9.19	9.07	9.03	9.01	9.00
20	13.61	10.84	10.36	10.23	10.18	10.15	10.14
22	15.00	12.18	11.63	11.49	11.44	11.41	11.40
24	16.60	13.50	13.01	12.87	12.81	12.76	12.75
26	18.25	15.17	14.62	14.46	14.39	14.34	14.32
28	20.20	16.88	16.27	16.09	16.01	15.96	15.94
30	22.51	18.81	18.13	17.92	17.84	17.78	17.76

TABLE III. Comparison of dispersion observed for adsorbed film and for liquid water.

Wave-length λ_v microns	$K\times10^7$	(n_v-1) relative values
0.3	1.032±0.095	0.995
0.5	0.932±0.017	0.932
0.7	0.904±0.031	0.915
0.9	0.893±0.041	0.907
∞	(0.876±0.058)	(0.898)

[11] *Handbook of Chemistry and Physics* (Chemical Rubber Publishing Company, Cleveland, 1947), thirtieth edition.

TABLE IV. Etalon temperature coefficients.

Etalon	Temperature change $\Delta\theta$°C	Temperature coefficient $\alpha\times10^6$/°C
15 mm etalon I	2.699	1.39±0.02
15 mm etalon II*	0.241	1.02±0.10
25 mm etalon	0.219	0.68±0.12

* Etalon II is the same as etalon I, but with different adjustment.

TABLE V. Thermal and elastic constants of materials.

Material	Linear temperature coefficient $\alpha\times10^6$ per °C	Linear compressibility $\tau/3\times10^{12}$ cm²/dyne
Fused quartz	0.42	0.89
'Invar'	0.9	0.28
Brass	18	0.37

where λ_v is the vacuum wave-length, $\Delta\epsilon$ is the change in the order of interference in going from an evacuated etalon to one in which the vapor density is ρ, t is the etalon spacing, n the refractive index of the vapor and K is the optical thickness of the adsorbed film for unit vapor density. Both $(n-1)$ and K are functions of wave-length, and if these be written in the Cauchy form, we may write for (1)

$$\lambda_v\Delta\epsilon/2t\rho=(n-1)/\rho+K/t$$
$$=A+B\lambda_v^{-2}+C\lambda_v^{-4}+\cdots \qquad (2)$$

where A, B, C, $\cdots$ are constants depending upon the etalon thickness. The values of $\lambda_v\Delta\epsilon/2t\rho$ were corrected for the compressibility of the spacer by the addition of a term $\tau h/3$ where τ is the volume compressibility of fused quartz and h the water vapor pressure. Using corresponding wave-length data, least-squares solutions of (2) were found for both etalon thicknesses:

(i) $t=14.983$ mm

$$\lambda_v\Delta\epsilon\times10^7/2t\rho=3.5968+0.02855\lambda_v^{-2} \qquad (3)$$

(ii) $t=24.940$ mm

$$\lambda_v\Delta\epsilon\times10^7/2t\rho=3.3635+0.02479\lambda_v^{-2} \qquad (4)$$

where λ_v is expressed in μ.

The standard deviations of the experimental points of (3) and (4) are 0.021 and 0.013, respectively. In terms of these quantities, the probable errors of the computed values range from about 0.06 at the lower wave-length limit (2000A) to about 0.003 at 5000A, and increase to about 0.01 at 10,000A. The probable error over the bulk of the range is less than 0.2 percent.

The expression for K was deduced from (3) and (4)

$$K\times10^7=0.876+0.0141\lambda_v^{-2}. \qquad (5)$$

The relation (5) was used to evaluate the individual values of the refractive index of water vapor, according to Eq. (2). The quantities thus derived were used to compute the least-squares equation

$$(n-1)\times10^7/\rho=3.0198+0.16365\lambda_v^{-2}+0.000133\lambda_v^{-4}. \qquad (6)$$

Eqs. (3), (4) and (6) are reproduced in graphical form in Fig. 3 where the individual data are also indicated.

COMPARISON WITH OTHER RESULTS AND WITH LIQUID WATER

Table I presents data from Eq. (6) along with computed probable errors and data from the work of the Cuthbertsons[1] and of Barrell and Sears.[2] Data for liquid water also appear, reduced to a comparable form. Bracketed values in the table are extrapolations beyond the range of the data from which the equations were originally derived.

The Cuthbertsons used the Jamin refractometer method of measurement, with vapor pressures greater than 760 mm and temperatures in the neighborhood of 140°C. The long light path used (36 to 37 cm) would prevent them from observing the effects of adsorbed films, if these occur at such high temperatures. Their data are generally higher and the dispersion is somewhat greater than those of the present experiments, indicating that the refractive index and hence the characteristics of the absorption bands depend somewhat on temperature.

Barrell and Sears used 20-cm Fabry-Perot etalons and moist air, but do not report any effects of adsorbed films. On the basis of the adsorption effects measured here, if their data be reduced by between 0.04 and 0.06 they are in good agreement with those of the present experiments.

It is noteworthy that at long wave-lengths the specific indices of all the data approach equality, the differences appearing at shorter wave-lengths. One surprising feature appears, in that the dispersion is affected more by temperature than by physical state. The index of liquid water is intermediate between the values of the present experiments and those of the Cuthbertsons.

CORRECTIONS TO THE REFRACTIVE INDEX OF AIR

The amount of water vapor present in air is usually determined in terms of the relative humidity, while the barometric pressure is the sum of the partial pressures of all the gases and vapors present. Using Eq. (6), the corrections for the presence of water vapor are somewhat unwieldy. Fortunately the corrections are not large and a table of values permitting interpolation of the corrections need not be too extensive under normal working conditions. The following equation is readily set up:

$$(n_m-1)_{h,\theta}=(n_a-1)_{h,\theta}-RF(\theta) \qquad (7)$$

where n_m and n_a are the refractive indices of moist air, and of dry air respectively at barometric pressure h and temperature θ. R is the relative humidity, and $F(\theta)$ is a factor depending only on the temperature,

$$F(\theta)=\frac{h_\theta}{760}(n_a-1)_{760,\theta}-\left(\frac{n-1}{\rho}\right)\rho_\theta. \qquad (8)$$

Here h_θ and ρ_θ are the pressure and density, respectively, of saturated water vapor at temperature θ, while $(n-1)/\rho$ is the specific index of water vapor as given by Eq. (6). The factors $F(\theta)$ for temperatures between 10°C and 30°C over the wave-length range 2000A to 10,000A appear in Table II. The refractive index of air n_a was obtained from the data of Meggers and Peters.

THE ADSORBED FILM OF WATER

The effective optical thickness of the film deposited on the mirror plates was given by the function $K\rho$ at low water vapor densities. If the effect of this film is primarily due to a water deposit, we would expect

$$K\rho=(n_w-1)s \tag{9}$$

where n_w is the refractive index of liquid water and s is the physical thickness of the film. Values of K and relative values of (n_w-1) reduced to correspond with K at a wave-length of 0.5μ appear in Table III.

The discrepancies between the values in Table III are considerably less than the probable errors in K which appear in the table. Using (9) with values at 0.5μ, we find for the thickness of the adsorbed film

$$s=29\rho\times10^{-8}\text{ cm}. \tag{10}$$

At a vapor density of 15 g/meter^3, the thickness of the film on aluminium oxide is about 420A and there are about 140 molecular layers. These figures are of the same order of magnitude as those for the films adsorbed on glass or quartz. No quantitative data have been found for the adsorption on aluminium oxide.

The use of silvered interferometer mirrors did not greatly reduce the effects of the adsorbed layer. Two sets of measurements using the 25 mm spacer and a vapor density of 17.779 g/meter^3 gave the average fractional change in optical path for eight neon lines ($\lambda\lambda$5854 to 6718A) as 3.501×10^{-7} meter^3/g, while the corresponding figure for the aluminized mirrors is estimated at 3.77×10^{-7} meter^3/g. The corresponding specific refractive index of water vapor is 3.062×10^{-7} meter^3/g.

MEASUREMENTS ON CARBON DIOXIDE

The data for the determination of the refractive index of carbon dioxide were obtained in a manner similar to those used for water vapor. Gas pressures between 10 and 20 mm of mercury were used for both the 15 and 25 mm etalon spacers. The refractive index was found to be directly proportional to pressure and there was no evidence of any adsorptive effect. The data were reduced to 20°C and the least-squares dispersion equation deduced

$$[(n-1)/h_c]\times10^{-7}=5.3749+0.032182\lambda^{-2}+0.000373\lambda^{-4}. \tag{11}$$

The probable error of points computed from (11) varies from 0.039 at 2000A to a minimum of 0.0028 at 4200A, increasing again to 0.0043 at 10,000A. The agreement with the data of other observers[3-5] who worked at pressures around an atmosphere is somewhat better than the probable error, indicating that there is no pressure effect exceeding the probable error of these experiments.

While carbon dioxide is present in normal atmospheric air only in very small amount, the amount found in laboratory air may be considerably greater. Following a procedure similar to that used in the case of water vapor, the correction to be added to the refractive index of air computed for the observed total pressure and temperature is given by

$$\Delta(n-1)\times10^7=h_c[(n_c-1)_{1,\theta}-(n_a-1)_{1,\theta}]=\frac{293h_c}{273+\theta}\{1.849+0.0306\lambda^{-2}+0.00033\lambda^{-4}\} \tag{12}$$

where h_c is the partial pressure of the carbon dioxide, n_c and n_a are the refractive indices of carbon dioxide and of air respectively at a pressure of one mm and temperature θ°C. The ideal gas temperature coefficient suffices for the purposes of these corrections where h_c is of the order of one mm and the temperature is close to 20°C.

TEMPERATURE COEFFICIENTS OF THE ETALONS

Advantage was taken of three occasions when the temperature control of the water bath broke down to obtain vacuum exposures at different temperatures, as the bath returned to its standard temperature. Data from these exposures permitted the computation of temperature coefficients for the two etalons. All coefficients listed in Table IV are the means of measurements at six or eight wave-lengths.

The linear temperature coefficients and linear compressibilities of some materials commonly used for the construction of Fabry-Perot etalons are listed in Table V.

It will be noted that the temperature coefficient of brass is considerably larger than that of fused quartz. Hence, rise in temperature has the effect of relieving the pressure of the springs holding the mirrors against the spacer, thus giving rise to an etalon temperature coefficient considerably larger than that predicted by the coefficient of fused quartz. Further, the effect should be greater for smaller etalon spacers. The data in Table IV are in accord with this analysis.

This effect suggests that the usual method of applying pressure corrections to refractive index data obtained using Fabry-Perot etalons of the general type used in these experiments may need some revision. The combination of fused quartz spacer and brass mounting would be expected to give a lower pressure coefficient than that based on fused quartz alone. Increase in pressure results in a greater relative decrease in the size of the spacer than in the mounting, and consequently the pressure of the springs is reduced, tending

to increase the etalon thickness. Thus the pressure corrections made to the present data are probably somewhat large. However, the compressibilities are more nearly equal than are the temperature coefficients and hence the pressure effect of the etalons is not expected to depart greatly from the compressibility of fused quartz. In the case of the present experiments the corrections applied for the pressure effect were not more than twice the probable error of the final dispersion equations. Thus the difference between the corrections actually applied and those based on the pressure effect of the etalon as a whole are probably considerably less than the probable error of the data of these experiments.

The pressure corrections were not of great importance in these experiments. However, they are of considerable importance where the refractive index of a gas at high pressure is being measured. For their work on the refractive index of air, Meggers and Peters used Invar spacers and brass mountings. The spacers were three Invar pins with hemispherical ends. In conjunction with the thick glass or quartz mirrors used, the effect would be to increase the etalon pressure coefficient relative to that of Invar. If the etalon coefficient were twice that of Invar, their determination of the refractive index of dry air would be increased by 5.6×10^{-7}.

Further, a comparison of the temperature coefficients in Table III indicates that the temperature effect is more marked for smaller etalon spacers. Meggers and Peters used etalon spacers ranging from 2 to 25 mm "depending on the source of light and the spectral region which was being used." Presumably this means that generally smaller etalon spacers were used in the ultraviolet. Hence, the dispersion equation obtained by Meggers and Peters would be expected to exhibit too low a dispersion in the violet and ultraviolet regions of the spectrum. Increases of the order of a few parts in 10^{-7} would bring the data of Meggers and Peters into good agreement with the data of Pérard and of Barrell and Sears in the visible region where the latter made their measurements. The fact that Meggers and Peters did not remove the carbon dioxide from the air used in their experiments would reduce their data by not more than about 1×10^{-7}.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 39, NUMBER 10 OCTOBER, 1949

Absorption Spectrum of Benzotrifluoride Vapor in the Near Ultraviolet*

H. SPONER AND D. S. LOWE**
Department of Physics, Duke University, Durham, North Carolina
(Received June 2, 1949)

The absorption spectrum of benzotrifluoride vapor at 2750–2400A was photographed in a medium Hilger quartz spectrograph. The band system can be interpreted in terms of an electronic transition A_1-B_1. The 0,0 band appears strongly at 37819 cm^{-1}. Several progressions of totally symmetric vibrations occur involving particularly frequencies of 750, 926, and 961 cm^{-1}. Other frequencies of 316, 538, 596, and 1031 cm^{-1} occur singly excited and in combination with the mentioned symmetric vibrations. Bands corresponding to the excitation of ground state frequencies up to 1336 cm^{-1} have been observed at higher pressures and found to coincide with Raman frequencies. The relation of the observed frequencies to modes of vibration is discussed. Spectral position and oscillator strength of the total electronic band is discussed in comparison to benzene, fluorobenzene and toluene.

INTRODUCTION

A STUDY of the near ultraviolet absorption of benzotrifluoride ($C_6H_5CF_3$, α-trifluorotoluene) was undertaken because it was expected that a comparison of the spectrum with the corresponding absorption regions of toluene and fluorobenzene would prove interesting and also would reveal additional information on assignments in these spectra. Furthermore, a comparison of the absorption strengths of the three molecules in the near ultraviolet should be of interest in connection with the different nature of the substituents in the three cases.

The near ultraviolet absorption spectrum of benzotrifluoride has not been reported before.[1,2] The Raman spectrum in the liquid state has been studied by Pendl and Radinger.[3] The infra-red spectrum was taken between 6 and 20μ in the liquid and vapor by Thompson and Temple.[4] Studies on the near ultraviolet fluorescence spectrum have been carried out by Mr. M. L. N. Sastri in this laboratory and will be published soon.

* This work was assisted by the ONR under contract N6ori-107, T.O.I., with Duke University. Part of this paper is based upon a Master's Thesis by D. S. Lowe, Duke University.

** Present address: Naval Research Laboratory, Washington, D. C.

[1] H. Sponer and D. S. Lowe, Phys. Rev. 74, 122A (1948).

[2] Dr. H. W. Thompson, Oxford, England informed the first author, H. Sponer, in a letter that the near ultraviolet absorption of benzotrifluoride has been studied in the vapor phase in his laboratory by Mr. Cave during 1947.

[3] E. Pendl and G. Radinger, Monats. f. Chem. 72, 382 (1939).

[4] H. W. Thompson and R. B. Temple, J. Chem. Soc. (1948), p. 1432.

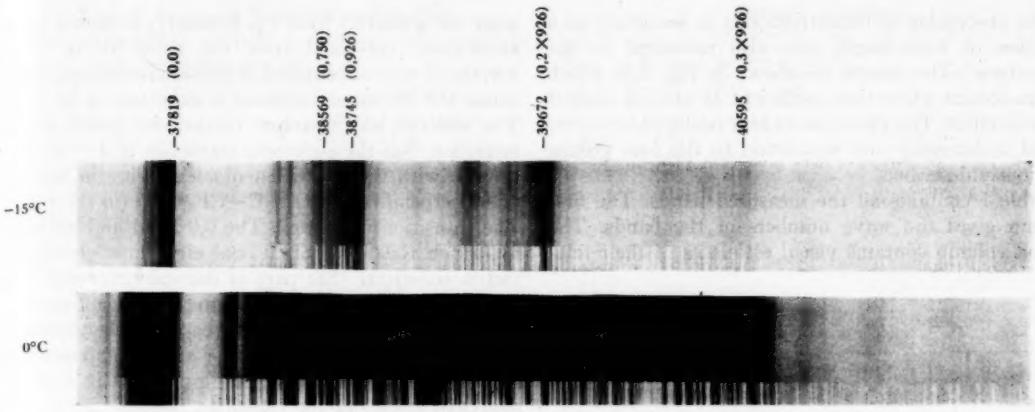

FIG. 1. Absorption spectrum of benzotrifluoride at different temperatures of the reservoir.

EXPERIMENTAL

The optical arrangement was the same as in previous similar studies from this laboratory. The light from a low voltage hydrogen lamp passed approximately parallel through the absorption cell and was focused onto the slit of a medium Hilger quartz spectrograph by means of two converging quartz lenses. Iron lines were used as the comparison spectrum. The spectrograms were taken on Eastman spectroscopic plates II-0.

A pure sample of benzotrifluoride was obtained through the courtesy of the Chemistry Department of Duke University. It was outgassed and distilled under vacuum into the side arm of an absorption tube of 25 cm length and 2.5 cm in diameter. Two separate furnaces regulated the temperatures of the main tube and the side arm, the latter always being at a slightly lower temperature than any other part of the tube. When pressures corresponding to temperatures below room temperature were desired, the furnace of the side arm was replaced by a Dewar flask containing acetone cooled by dry ice. The absorption was studied in dependence of the pressure of the substance, the pressures being varied between 0.3 and 80 mm Hg. Benzotrifluoride has a boiling point[5] of 102.30°C and a melting point[6] of −29.05°C.

Most spectrograms were taken at temperatures between −20°C and 0°C corresponding to a pressure range of about 2–10 mm Hg. Exposure times were usually 4 minutes with Eastman spectroscopic plates II-0. A few spectrograms were taken on the large Hilger quartz spectrograph *E*1 but it was found that the band heads were not sharp enough to gain by the greater dispersion an improved measurement of the bands. The plates were measured on the Hilger comparator to, in general, a few wave numbers while the more diffuse bands could be measured only to about five wave numbers.

[5] J. H. Saylor and F. Field, J. Am. Chem. Soc. **68**, 2649 (1946)
[6] Booth, Elsey, and Burchfield, J. Am. Chem. Soc. **57**, 2066 (1935).

EXPERIMENTAL RESULTS

The near ultraviolet absorption spectrum of benzotrifluoride lies in the region of 2750−2400A. Fig. 1 shows the spectrum at the reservoir temperatures −15°C and 0°C according to vapor pressures of 3.3 and 10 mm. The strongest bands of the spectrum may be clearly recognized when the reservoir is at −40°C corresponding to a pressure of 0.3 mm Hg. The spectrum is best developed when the temperature of the reservoir is −15°C (pressure of 3.3 mm). At these lower pressures the spectrum consists of four groups of bands, the last of which, on the short wave-length side, is of much weaker intensity than the others. Even a rough comparison of the four groups indicates that they are not mere repetitions of the first group, but that the groups of shorter wave-lengths are more complex than the first group although they show structural similarities to the first group. Although the bands are diffuse, their degradation to the red may be clearly seen. The bands broaden with increasing pressure and eventually merge into each other. When the temperature of the reservoir reaches 5°C corresponding to a pressure of about 12 mm Hg, the absorption becomes continuous except for a few bands which now appear on the long wave-length side of the spectrum.

Benzotrifluoride has another absorption region on the short wave-length side of the spectrum under discussion. Spectrograms taken on ultraviolet sensitized plates showed no structure in this second region, the maximum intensity of which lies in a range beyond the reach of our experimental set-up. When the temperature of the reservoir is −10°C, this absorption extends to about 2200A. With increasing pressures it spreads toward longer wave-lengths while the absorption spectrum under discussion broadens towards shorter wave-lengths. This process can easily be followed when using Eastman 40 plates. At a reservoir temperature of 40°C (80 mm pressure), there is left only a weak transmission region of about 150A between the two absorption regions.

The absorption of benzotrifluoride in iso-octane as a function of wave-length was also measured in our laboratory.[7] The results are shown in Fig. 2 in which the molecular extinction coefficient is plotted against wave number. The curve shows four peaks which correspond in intensity and separation to the four groups previously described.

Table I contains all the measured bands. The first column gives the wave numbers of the bands. The second column contains visual estimates of their relative intensities. This column is divided into subcolumns such that all bands, whose intensities are estimated from plates taken with the reservoir at a given temperature are in the same sub-column. In order to give an idea of the relation between the different intensity scales, the intensities of some bands are given at more than one temperature. The following notations have been used: *es*=extremely strong, *vs*=very strong, *s*=strong, *s–m*=strong to medium, *m*=medium, *m–w*=medium to weak, *w*=weak, *vw*=very weak, *ew*=extremely weak. The last column will be discussed in the next section.

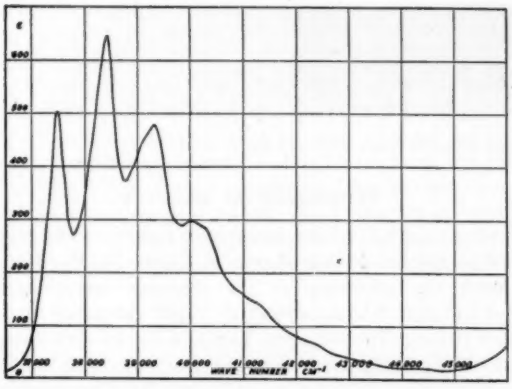

FIG. 2. Absorption of benzotrifluoride in iso-octane solution.

DISCUSSION AND ANALYSIS OF THE SPECTRUM

The symmetry properties of benzotrifluoride are the same as those of toluene. Pitzer and Scott[8] assigned C_{2v} symmetry to toluene in their discussion of the Raman spectrum. In doing so, they assumed that because of the small restricting potential against free rotation the CH_3 group may be considered as a one-atom substitution in its effect on the benzene ring. It is to be expected that the CF_3 group has a larger potential of hindrance and that hence the assignment of C_{2v} symmetry to benzotrifluoride is justified to a lesser degree. However, since the departure from C_{2v} symmetry is caused by the substituent only and since the main bands of the spectra of monosubstituted benzenes involve ring vibrations, the discussed influence is expected to be small. The analysis was therefore carried out under the assumption that the electronic transition is A_1-B_1 as in toluene with the transition moment lying in the ring plane perpendicular to the $C-CF_3$ bond (in the x axis). The transition is allowed. The 0,0 band and all bands for which the total vibrational eigenfunction is symmetric constitute that part of the spectrum which results from the x component of the electric moment. It represents the majority of the bands. Other bands in the spectrum may contain the y and z components (y axis through the C of the CF_3 group, z axis $\perp$ to the molecular plane) of the transition moment if the total vibrational eigenfunction associated with the electronic transition has the symmetry B_1 or A_2. Experience has shown that these bands are in general of weaker intensity than the main bands. This is particularly true for symmetry A_2 in which case the transition moment is perpendicular to the ring plane (z-component).

The band 37819 cm^{-1} on the short wave-length side of the first group is taken as the 0,0 band of the spectrum. Bands on the long wave-length side of the 0,0 band arise from transitions from vibrational levels in the ground state, while the main part of the spectrum, which lies on the short wave-length side of the 0,0 band, arises from transitions from the vibrationless ground state to vibrational levels in the upper electronic state. In representing the results of the analysis the following notation is used for the assignment column in Table I. The number at the left of the comma indicates the wave number of the vibration in the ground electronic state. The numbers at the right of the comma, unless preceded by a minus sign, give the wave numbers of the vibrations in the excited electronic state. The "negative" numbers refer to difference frequencies of 1,1 transitions.

The frequency values of vibrations as found to occur in the ground and excited electronic states of the spectrum under consideration are given in Table II along with the Raman and infra-red frequencies.[3,4] The characteristic structure of many bands aided in the correlation of frequencies. For example, at a pressure of 4.6 mm Hg (reservoir temperature $-10°C$) the band assigned 338,0 looks very much like the one assigned 0,316. Thus it was concluded that the vibrational frequency of 338 cm^{-1} in the ground state corresponds to the frequency of 316 cm^{-1} in the excited state. Bands which involve the 316 frequency besides some other vibrational frequencies have a similar structure.

Before examining the structure within a group the appearance of the different groups will be briefly discussed. Starting with the first group and taking the 0,0 band 37819 as representative, this group is repeated at intervals of 926 cm^{-1} at 38745, 39672, and 40595 cm^{-1}. The first group also repeats itself on the long wave-length side of the 0,0 band at 36815 cm^{-1}. The separa-

[7] H. P. Stephenson, Master's Thesis, Duke University, 1949. See also E. J. Hackney, Ph.D. Thesis, Chemistry Dept., Duke University, 1949.

[8] K. S. Pitzer and D. W. Scott, J. Am. Chem. Soc. 65, 803 (1943).

TABLE I. Absorption bands of benzotrifluoride.

Wave-lengths Å	Wave numbers cm⁻¹	Intensities at different reservoir temperatures 39°C	14°C	5°C	Assignments
2740.2	36483	*ew*			1336,0; 1004+338,0
26.0	673	*vw*			1004+141,0
23.4	708	*vw*			1111,0; 1004,−75−36
21.1	739	*vw*			1004,−75
18.4	776	*m–w*	*ew*		1004,−36
15.5	815	*m*	*vw*		1004,0
2692.8	37125		*ew*		623,−75; 662,−36
90.5	157		*vw*		662,0; 623,−36
87.8	196		*s–m*	*ew*	623,0
77.1	343			*vw*	141+338,0
75.2	369			*vw*	338,−75−36
72.8	403			*vw*	338,−75
		5°C	−5°C	−20°C	
2669.9	37444	*s*	*vw*		338,−36
67.2	481	*vs*	*w*	*ew*	338,0
61.3	564		*vw*		141,−75−36
58.6	603		*vw*	*ew*	141,−75
55.8	642		*s–m*	*vw*	141,−36
53.3	678		*vs*	*m–w*	141,0
51.1	709			*m–w*	0,−75−36
48.8	743			*m–w*	0,−75
45.9	783			*s*	0,−36
43.4	819		*es*	*vs*	0,0
40.2	863		*ew*		
37.6	902		*ew*		
29.2	38023		*ew*		0,316−75−36
26.8	058		*ew*		0,316−75
24.0	098		*m*	*ew*	0,316−36
21.7	135		*s–m*	*w*	0,316
15.9	216		*w*		141,538
13.7	248		*vw*		0,538−75−36
11.4	282		*s–m*	*ew*	0,538−75
08.7	322		*s*	*vw*	0,538−36
06.3	357		*vs*	*m*	0,538
02.4	415			*m*	0,596
		−10°C	−20°C		
2594.4	38533			*w*	0,750−36
92.0	569			*m*	0,750
89.7	603			*ew*	141,926
87.3	639			*w*	0,926−75−36
85.0	673			*w*	0,926−75
82.7	708			*s–m*	0,926−36; 0,961−75
80.2	745			*s*	0,926
77.9	780			*m*	0,961
75.7	813	*vw*			0,1031−36
73.2	850	*m*		*w*	0,1031
70.8	887	*ew*			0,750+316
59.2	39063	*m–w*		*vw*	0,926+316
57.1	095	*vw*		*ew*	0,961+316
52.3	169	*w*		*ew*	0,750+596
49.8	207	*vw*			0,926+538−75
47.3	246	*m*			0,926+538−36
44.8	284	*s*		*vw*	0,926+538
42.9	313	*s*		*vw*	0,961+538; 0,2×750
41.1	341	*s–m*		*vw*	0,926+596
38.1	388	*m*		*ew*	0,538+1031
36.0	420	*ew*			0,926+750−75
33.6	458	*w*			0,926+750−36
31.1	497	*s–m*		*vw*	0,926+750
28.8	533	*m*		*vw*	0,961+750
26.6	567	*w*			0,750+1031−36
24.4	602	*m*			0,2×926−75; 0,1031+750
22.2	636	*s*			0,2×926−36
19.9	672	*vs*			0,2×926; 0,2×961−75
17.7	707	*s*			0,926+961; 0,2×961−36
15.5	742	*m*			0,2×961
13.1	780	*m*			0,1031+926
11.1	811	*w*			0,1031+961
2499.8	991	*vw*			0,2×926+316
97.7	40025	*ew*			0,926+961+316
95.6	058	*ew*			0,2×961+316; 0,3×750

TABLE I.—*Continued.*

Wave-lengths Å	Wave numbers cm⁻¹	Intensities at different reservoir temperatures −10°C	0°C	Assignments
93.2	097	*ew*		0,2×926+538−75−36
91.0	132	*ew*		0,2×926+538−75
88.7	169	*ew*		0,2×926+538−36
86.5	205	*w*		0,2×926+538
84.4	239	*m–w*		0,926+961+538; 0,926+2×750
82.4	271	*w*		0,2×926+596; 0,2×961+538
79.7	315	*vw*		0,926+1031+538
77.7	348	*ew*		0,2×926+750−75
75.6	382	*ew*		0,2×926+750−36
73.2	421	*vw*		0,2×926+750
71.0	457	*vw*	*s*	0,961+926+750
68.9	492	*ew*	*m*	0,2×961+750
66.7	528	*vw*	*s*	0,926+1031+750
64.8	559	*vw*	*vs*	0,3×926−36
62.6	595	*w*	*vs*	0,3×926
60.4	632	*w*	*vs*	0,2×926+961
58.4	665	*vw*	*s*	0,926+2×961
56.3	701	*ew*	*m*	0,3×961
53.9	739	*ew*	*m*	0,926+961+1031
52.1	769		*vw*	0,2×961+1031
44.1	902		*w*	
40.0	971		*vw*	
37.2	41018		*ew*	0,3×961+316
35.0	055		*ew*	0,926+961+1031+316
30.5	131		*ew*	0,3×926+538
28.6	163		*ew*	0,3×926+596−36
26.7	196		*ew*	0,3×926+596
24.4	235		*ew*	0,2×926+961+596
22.3	271		*vw*	0,926+961+1031+538; 0,3×926+750−75
20.3	305		*vw*	0,3×926+750−36
17.9	346		*vw*	0,3×926+750
15.8	382		*w*	0,2×926+961+750
13.7	418		*vw*	0,926+2×961+750
11.8	450		*vw*	0,3×961+750
09.6	488		*vw*	0,4×926−36
07.9	517		*vw*	0,4×926
05.6	557		*vw*	0,3×926+961
03.6	592		*vw*	0,2×926+2×961
01.7	624		*ew*	0,926+3×961

tion from the 0,0 band is 1004 cm⁻¹. By comparison with interpretations given for corresponding bands in other electronic spectra of monoderivatives, this value of 1004 cm⁻¹ corresponds to the excitation of a totally symmetric carbon ring vibration in the ground electronic state (carbon breathing vibration). It is assumed that the 926 cm⁻¹ corresponds to the same vibration in the excited state. Another, less distinct repetition of the first group within the second group occurs at 38569 cm⁻¹ giving a separation of 750 cm⁻¹ from the 0,0 band. This frequency may belong to the C−CF_3 valence vibration. The two last bands on the short wave-length side of the second group are 961 and 1031 cm⁻¹ apart from the 0,0 band. It is reasonable to assign the 961 to another symmetric carbon vibration expected in this frequency region. No suggestion will be made regarding the assignment of the 1031 cm⁻¹ frequency.

On the long wave-length side the second group contains two fairly strong bands at 38357 and 38415 giving the separations 538 and 596 cm⁻¹ from the 0,0 band. It is plausible to correlate these frequencies in the

TABLE II. Wave numbers of some vibrations of benzotrifluoride occurring in ultraviolet absorption.

Raman (ref. 3)	Infra-red (ref. 4)	Ultraviolet Absorption ground state	Ultraviolet Absorption excited state	Suggested symmetry
139 (7)		141		β_1
339 (5)		338	316	α_1?
618 (3)		623	538	β_1
656 (½)	664 *s*	662	596	α_1
771 (8)	771 *s*		750	α_1
1005 (9)	1005	1004	926	α_1
1027 (4)	1027 *s*		961	α_1
1068 (0)	1065		? 1031	
	1080 *vs*			
1116 (0)		1111		
1324 (3)	1325 *vs*	1336		

excited state to the Raman frequencies 618 and 656 in the ground state both of which have been observed in ultraviolet absorption (Table II). A comparison of the benzotrifluoride spectrum with that of toluene suggests that the mentioned two bands correspond to the toluene bands at 37934 and 38005 with separations of 456 and 528 wave numbers from the 0,0 band. The two toluene bands were explained[9] as arising from transitions from the vibrationless ground state to the oscillating excited state in which either component vibration is present resulting from the removal of the degeneracy of the 606 cm^{-1} ϵ^+g vibration in benzene upon substitution of the methyl group. The frequencies 456 and 528 were correlated with the ground state frequencies 521 and 623 known from Raman spectra. The lower value belongs to the α_1 component according to the stronger participation of the substituent in this mode of vibration than in the β_1 mode which is found practically independent of the substituent. If carrying through the analogy in benzotrifluoride the frequencies 538 and 596 should be assigned to the corresponding α_1 and β_1 components and be correlated with the Raman lines 618 and 656. That the α_1 component should have a larger value than that in toluene is surprising as is the increase in the β_1 component. Polarization measurements of the Raman lines, which would aid in the identification, are missing. We consider it as more likely that the 538 (and 618) frequency corresponds to the non-totally symmetric component β_1 and that the 596 (and 656) frequency belongs to a different type of vibration. As for the numerical value it could represent the $C-CF_3$ valence vibration which should have a smaller wave number than the $C-CH_3$ vibration in toluene. The latter is 786 in the ground and 751 in the excited state. Now it was suggested in the last paragraph that the 750 frequency in the excited state may represent this $C-CF_3$ vibration. Its value in the ground state is taken as 771 cm^{-1}. Since this type of totally symmetric vibration in monosubstituted benzenes always gives rise to a strong Raman line and since the 656 is a weak line, the interpretation given in the last paragraph has our preference and the 656 frequency must be assigned to a yet unidentified mode of vibration. The vibration occurs strongly in the infra-red with a value of 664 cm^{-1}. As may be seen from Tables I and II our value is 662 cm^{-1}. If the 0,596 band really represents this vibration in the upper electronic state, then the intensity of the band suggests symmetry α_1 or β_1 for this vibration. It is unlikely that a carbon vibration of these symmetry types occurs in the considered frequency region. If it were a hydrogen vibration, bands should be found in corresponding places in other monosubstituted benzene spectra. Now there are bands found in fluorobenzene (38597, 38581) and in toluene (38260/63) which very probably involve a hydrogen vibration of the same mode. The respective separations from the 0,0 bands are 779 (762) and 783 cm^{-1}, and the corresponding values in the ground state are 829 and 842 cm^{-1}. But it is more plausible that the latter values correspond to the weak Raman line at 839 cm^{-1} in benzotrifluoride than to the 656 line. The question then arises whether the 656 frequency may represent a vibration of the CF_3 group. The Raman work on fluoroform[10] (HCF_3) was consulted for information. It was found that a fairly strong Raman line has been reported at 697 cm^{-1} and was assigned to the symmetric bending vibration of the CF_3 group. The value may very well drop to 656 (or 664) in benzotrifluoride and it is plausible that it will give rise to a weak Raman line in this substance. If the assignment of the 656 in the ground state and the 596 in the upper electronic state to a CF_3 bending vibration be correct, the question may be raised why this vibration occurs so strongly excited with the electronic transition. As expected and discussed carbon ring vibrations will be more influenced by the electronic excitation than hydrogen vibrations and vibrations of the side chain. This is confirmed by the experiments. It is true that the 596 occurs in fewer combinations than the 538 frequency, and it does not form a progression. It occurs coupled mostly with the carbon "breathing" vibration and the $C-CF_3$ valence vibration, which is reasonable. No final explanation can be given for its relatively strong appearance. It may be remembered in this connection that in the absorption spectrum of methyl iodide[11] at 2100–1500A there has been observed the symmetric CH_3 deformation vibration in the upper (1090 cm^{-1}) and in the lower (1237 cm^{-1}) electronic state although this absorption system is due to an excitation of an iodine electron. It is possible that this vibration occurs also in the toluene spectrum. Its value in the ground state should be about 1300–1400. There are three unassigned bands[9] at 38672, 38705, 38739 cm^{-1} with separations of 1195, 1228, 1262 wave numbers from the 0,0 band, any one of which could correspond to the CH_3 deformation vibration in the excited electronic state. Combinations with the totally symmetric

[9] Ginsburg, Robertson, and Matsen, J. Chem. Phys. 14, 511 (1946).

[10] G. Glockler and W. F. Edgell, J. Chem. Phys. 9, 224 (1941).

[11] A. Henrici and H. Grieneisen, Zeits. f. physik. Chemie B30, 1 (1935); see also discussion in H. Sponer and E. Teller, Rev. Mod. Phys. 13, 111 (1941).

carbon ring vibrations and the C—CH_3 valence vibration fall mostly on top of other bands. In benzotrichloride the CCl_3 bending vibration may be expected to have a value of about 350 or less in the ground state and about 300 or a little less in the excited state. Actually a maximum has been found at that position in benzotrichloride vapor[12] but it is doubtful that it should be explained entirely in this way, instead it is rather expected that this maximum contains more than one band, one of which may correspond to the CCl_3 vibration.

If the suggested tentative interpretation of the 38415 band in benzotrifluoride should be correct then the toluene band at 37934 has no obvious analogue in benzotrifluoride, i.e., the α_1 vibration resulting from the splitting of the 606 benzene vibration is either absent or has a considerably lower value in benzotrifluoride than in toluene. Now a fairly strong Raman line was found[2] at 339 cm^{-1}. The vibration occurs in absorption and was correlated before with the vibrational frequency of 316 cm^{-1} in the upper state. Although it is not impossible it seems not very plausible to assign the frequency of 339 cm^{-1} to this α_1 vibration because a somewhat higher value would look more reasonable. But there is no line in this region available in the rather incompletely known Raman spectrum of benzotrifluoride which could account for this α_1 vibration, although its appearance may be expected. It is believed that a study of the vapor spectra of benzylfluoride and benzalfluoride would contribute to the solution of these tentative assignments. One then could follow the development of the spectrum from toluene over these two compounds to that of benzotrifluoride. Even though Raman measurements of the two incompletely fluorinated toluenes are missing, the comparison would be worthwhile.

Very little need be said about the other groups on the short wave-length side of the 0,0 band. Almost all bands in this region can be interpreted as combinations of the already discussed vibrational frequencies in the upper electronic state (see Table I).

We turn next to the discussion of the pattern of the first group which is, as mentioned before, repeated in the subsequent groups. The group consists of five distinct bands with the 0,0 band and the band adjacent to it being the strongest. The separations of the five bands from the 0,0 band are 36, 75, 110, and 141 cm^{-1}. The band 0,−36 is almost as strong as the 0,0 band. It is therefore concluded, that the difference of 36 cm^{-1} represents the 1,1 transition of a low vibration. The numerical figure varies from 34 to 39 cm^{-1}, but 36 cm^{-1} has been taken as representative value in Table I. The occurrence of a pair of strong bands at the beginning of a group is characteristic of many substituted benzene spectra. In fluorobenzene the distance between them is 67 cm^{-1}, in chlorobenzene 60 cm^{-1}, in toluene 59 cm^{-1}, to mention only a few examples of monoderivatives. In no case was it possible to give an unambiguous interpretation of these separations. Although only a careful study of the temperature dependence of these bands can decide the frequency region of the vibration involved it seems safe to assume that it will be about 200–400 cm^{-1}. The vibration may correspond to the lowest observed Raman line. This line has a fairly high intensity and, where measured, is depolarized. It may represent the planar bending vibration of the substituent. In benzotrifluoride, the Raman line is at 139 cm^{-1}. The fifth band of the first group occurs 141 cm^{-1} displaced from the 0,0 band and this is believed to correspond to the 139 Raman line. The vibration is likely of the β_1-type. The band would then correspond to a transition having its moment in the y-axis (through the C atom of the CF_3 group) just like the band 0,538 on the violet side of the 0,0 band. If the 139 frequency should be involved in the 1,1 transition with a drop of 36 cm^{-1}, its value in the excited electronic state would be 139−36=103 cm^{-1}. No band has been observed at this position.

The third band of the group has a much lower intensity than the first two, hence the separation of 75 cm^{-1} should correspond to a 1,1 transition of a vibration with higher frequency than the 139 just discussed. The differences 623−538=85 (or 80 cm^{-1} if using the Raman value 618) and 662−596=66 are considered to be too far off from the observed separation of 70–78 cm^{-1} (representative value of 75 cm^{-1} in Table I), to be identified with this separation. No suggestion can be put forward as to what vibration may be involved in the corresponding 1,1 transition. It cannot be excluded that the separation of 75 cm^{-1}, which shows the wide variation from 70 to 78 cm^{-1}, is a 2×36 cm^{-1}, or at least partly so, and the 0,−76−36 may contain a 0, −3×36 cm^{-1}. The two bands have the same intensities but are much weaker than the 0,0 and 0,−36 bands which are of comparable intensity.

DISCUSSION OF SPECTRAL POSITION AND OSCILLATOR STRENGTH OF THE TOTAL ELECTRONIC BAND IN COMPARISON TO BENZENE, FLUOROBENZENE, AND TOLUENE.

As is known and has been discussed by several authors the wave-length region of the near ultraviolet absorption spectra of substituted benzenes is shifted toward longer wave-lengths in comparison to the benzene spectrum. In comparative discussions of solution spectra the entire wave-length regions or the spectral positions of the maximum peaks are considered usually. If the spectra arise from allowed transitions, i.e., if the bearer-molecules are not highly symmetric, the peaks can be compared directly. But if a comparison with the benzene spectrum is included, proper adjustment has to be made to the fact that benzene has a 0,1 and 1,0 transition of a particular vibration instead of the 0,0 band, and that the 0,0 band is a calculated quantity only. Table III contains the 0,0 bands of the four molecules under comparison. From the table it is apparent that the red

[12] W. F. Hamner and F. A. Matsen, J. Am. Chem. Soc. 70, 2482 (1948).

TABLE III. 0,0 bands and their shifts with respect to benzene.

Molecule	0,0 band in cm^{-1}	Shift	$f \cdot 10^3$
benzene	38089 (calc.)	—	1.6
fluorobenzene	37819	−270	8.9*
toluene	37477	−612	3.1*
benzotrifluoride	37819	−270	7.1*
chlorobenzene	37052	−1037	~3.0
benzylchloride	37110	−979	
benzalchloride	37020	−1069	
benzotrichloride	36740	−1349	

* See reference 7.

shift in benzotrifluoride, instead of being larger than in toluene, is the same as in fluorobenzene. Herzfeld[13] has used the molecular orbital method to explain and approximately calculate the wave-length shift of the 0,0 transition caused by an ortho-para directing substituent. He finds that the lowering of the energy of the electronic transition results from the fact that, upon excitation, a migrating electron leaves the excited orbital and migrates into a previously filled, lower orbital of the resonating system. While F and the group CH_3 are ortho-para directing (migration into the ring), the group CF_3 is meta directing (migration out of the ring). In this case Herzfeld's calculation means that the migrating electron is shared between the resonating group and the unoccupied orbitals of the substituent. It so happens that the electron migration from the fluorine into the ring in fluorobenzene has the same effect on the electronic transition as has the electron migration out of the ring into the CF_3 group. Experimentally, it has been found that all substitutions produce a red shift. Only recently Hamner and Matsen[12] studied the series benzylchloride, benzalchloride, benzotrichloride in the vapor phase and concluded that the migration of electrons out of the ring has qualitatively the same effect on the spectra as the migration into the ring. The series presents a gradual transition from ortho-para to meta direction. The positions of the 0,0 bands and the shifts have been added in Table III together with the corresponding data for chlorobenzene. In comparison to toluene the successive replacement of the methyl hydrogens goes hand in hand with a lowering of the electronic transition, with the second step showing the smallest successive change. It is interesting that the pure electronic transitions of benzyl- and benzalchloride have about the same energy as that of chlorobenzene.

We compare next the intensity of the near ultraviolet absorption system of benzotrifluoride with the corresponding intensities in fluorobenzene, toluene, and benzene. It was expected[13] that in general a large red shift will go with a strong intensification over the benzene transition. However, there are so many exceptions to this assumption that its general validity seems doubtful. Chlorobenzene (and bromobenzene) with a medium shift absorbs only twice as strongly as benzene itself while fluorobenzene with a small shift absorbs 6 times stronger. Toluene with a medium shift has an absorption intensity of only twice that of benzene, and benzotrifluoride with a small shift shows a factor of intensification between 4 and 5. This may be seen from Table III where the last column contains the oscillator strengths f for the four molecules compared. Sklar[14] has discussed the intensification of several monoderivatives in terms of the migration effect. Because of lack of data he assumed fluorobenzene to be a weak absorber like toluene, and he attributed the weakness to the high electronegativity of the fluorine atom which would make the ionization potential of the radical large. However, it is believed here that the relatively small C−F distance (1.34A) and the fact that fluorine has two unbound p electrons available for migration offsets the other factor. That benzotrifluoride has a larger absorption intensity than toluene may result from the large inductive effect of the meta directing CF_3 group which facilitates electronic migration out of the ring. In toluene the migration into the ring is due to bound electrons from the C−H bonds, producing a small effect.

Since large absorption intensity is usually found in monoderivatives with substituents of large directing power it is not surprising that the strongly meta directing CF_3 group[15] produces also a noticeable intensification. Similarly, fluorine should have a higher ortho-para substitution rate than the other halogens. This is found to be the case.[16]

[13] K. F. Herzfeld, Chem. Rev. 41, 233 (1947).

[14] A. L. Sklar, J. Chem. Phys. 7, 984 (1939).

[15] F. Swarts, Bull. Acad. Belgique (1920), p. 395.

[16] M. L. Bird and C. K. Ingold, J. Chem. Soc. (1938), p. 918.

The Aberrations of Slightly Decentered Optical Systems*

L. Ivan Epstein
Scientific Bureau, Bausch and Lomb Optical Company, Rochester, New York
(Received June 2, 1949)

Two separate cases of aberrations due to decentration are discussed: In one, the field of view is small, and the aperture and spherical aberration are finite. In the other, the aperture is small, and the field of view, astigmatism, curvature of field, and distortion are finite. In either case, only small decentrations and tilts are admitted, and small quantities of order higher than the first are neglected. Since each member of the decentered system is itself a coaxial system, skew rays may be traced through each member with the aid of the laws of orthogonal systems. It is found that decentrations in a wide-aperture system give rise to coma, whereas in a wide-field system they give rise to a kind of distortion. A numerical example of each is given, and the resulting aberrations are compared to their third-order approximations.

WIDE-APERTURE SYSTEMS

THE third-order aberrations of a slightly decentered optical system consisting of spherical surfaces only have been investigated by A. E. Conrady.[1] E. H. Linfoot[2] has extended Conrady's method to the case where some of the surfaces are aspheric. Methods of ray-tracing through decentered optical systems have been given by A. Thomscheidt[3] and F. Temmerman.[4] Recently, A. Maréchal[5] has expressed the coma due to decentration of a wide-aperture system in terms of offense against the Staeble-Lihotzky condition. In another paper,[6] Maréchal has given a new treatment of the third-order aberrations caused by decentration, as well as a theory of higher order aberrations, where the individual aberrations are represented as the terms in a series expansion. The present method, while based on ray-tracing, permits a representation of one of the aberrations (usually the most objectionable in practice) for arbitrary decentrations.

Consider first of all an optical system having an infinitesimal field of view. The directions of the decentrations and tilts of the members of the system are left arbitrary, but their magnitudes are first-order infinitesimals. By a member we shall understand any surface of revolution, or any coaxial sequence of such surfaces, which are decentered and tilted as one body. No limitations are placed on aperture or spherical aberration. All infinitesimals of order higher than the first are neglected.

The individual members of our system are themselves perfectly coaxial optical systems. Consider first of all the system with all of its members in perfect alignment. A meridional ray is traced through the system from the axial point of the object plane. This at once defines an entire conical shell of rays. Now if the object point moves off axis, or if some of the members of the system are decentered or tilted, each ray changes its position and orientation by infinitesimal amounts.

The portion of a coaxial optical system lying in the immediate vicinity of a meridional ray forms what is called an orthogonal system.[7,8] The laws of such orthogonal systems are quite simple. First of all, there is Lippich's theorem: If a skew ray deviates from a given meridional ray by no more than an infinitesimal at any point, then its projection upon the tangential plane in the image space is the image of its projection upon the tangential plane in the object space, and its projection upon the sagittal plane in the image space is the image of its projection upon the sagittal plane in the object space. Figure 1 shows a portion of the axis of the νth member, the original meridional ray in the object space of this member, and the tangential and sagittal planes through it. The projections of a typical skew ray upon these two planes are also shown. These projections may be traced through the νth member by elementary

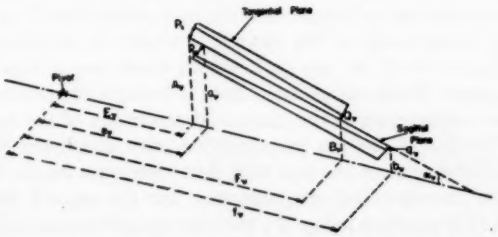

FIG. 1. A typical skew ray and its projections in the object space of the νth member.

* Since the first part of this paper was presented at the meeting of the Optical Society of America in New York, New York, 1949, the recent papers on the same subject by A. Maréchal (see references 5 and 6) have come to the author's attention. But, since Dr. Maréchal's method of solving this problem is quite different from ours, it is believed that the present paper is still of interest.

Note added in proof.—Another paper by W. G. Stephan (App. Sci. Research **B1**, 273 (1949)), which has just come to the author's attention, contains the same subject matter as reference 6, which it antedates, and goes into more detail.

[1] A. E. Conrady, M. N. R. A. S. **79**, 384 (March 1919).
[2] E. H. Linfoot, Proc. Phys. Soc. London **58**, 65 (1946).
[3] A. Thomscheidt, Zeits. f. Instrumentenk. **61**, 201 (1941).
[4] F. Temmerman, Rev. d'Optique **26**, 423 (1947).
[5] A. Maréchal, Rev. d'Optique **27**, 649 (1948). At the time of this writing, the second part of this paper has not yet appeared in print.
[6] A. Maréchal, Comptes Rendus **228**, 668–670 (1949). The measure of decentration and tilt used by us was first proposed in this paper.
[7] See M. Herzberger, *Strahlenoptik* (Verlag. Julius Springer, Berlin, 1931), Sections 22–25, or reference 8.
[8] K. Luneberg, *Mathematical Theory of Optics* (Brown University Lectures, 1944), Chapter IV, Section 35.

TABLE I. Linear and circular coma per unit decentration and tilt in each space, for two rays, together with (the lower of each pair of entries) the third-order approximation.*

Radius	+25.9	−253.0	−60.0	+23.6	D	−144.0	+24.9	−39.4	
spacing	4.1	2.9	1.7	4.6	3.3	1.7	5.1		
index	1.6088	1	1.5602	1	1	1.5602	1.6165		
									Diaphragm
zonal ray, $K=8$									
decentration									
linear	−0.1506	−0.1690	−0.0630		+0.1034	+0.1184	+0.0912		+0.009
	−0.1431	−0.1597	−0.0674		+0.0791	+0.1090	+0.0858		+0.024
circular	−0.0766	−0.0861	−0.0306		+0.0562	+0.0609	+0.0464		+0.002
	−0.0716	−0.0798	−0.0337		+0.0396	+0.0545	+0.0429		+0.012
tilt									
linear	−3.35	+1.798	−4.52		−1.414	−2.45	−3.59		−0.03
	−3.33	+1.323	−4.15		−1.318	−2.37	−3.38		−0.08
circular	−1.677	+0.985	−2.33		−0.722	−1.237	−1.832		−0.00
	−1.666	+0.662	−2.08		−0.659	−1.184	−1.691		−0.04
marginal ray, $K=11$									
decentration									
linear	−0.298	−0.336	−0.1054		+0.258	+0.243	+0.1817		−0.025
	−0.271	−0.302	−0.1275		+0.1495	+0.206	+0.1623		+0.046
circular	−0.1533	−0.1740	−0.0478		+0.1508	+0.1286	+0.0942		−0.028
	−0.1353	−0.1510	−0.0638		+0.0748	+0.1030	+0.0811		+0.023
tilt									
linear	−6.31	+4.44	−9.38		−2.84	−4.70	−7.16		+0.08
	−6.30	+2.50	−7.85		−2.49	−4.48	−6.39		−0.15
circular	−3.14	+2.58	−4.98		−1.482	−2.38	−3.72		+0.07
	−3.15	+1.25	−3.92		−1.246	−2.24	−3.20		−0.08

* U. S. Patent No. 1558073/1922, Bausch and Lomb Optical Company.

methods, and so we obtain the projections of the skew ray in the image space of the νth member, and hence, by implication, the skew ray itself. However, for the sake of greater generality in tracing an entire conical shell of rays, we shall do well to make use of another property of orthogonal systems: A line element perpendicular to the original meridional ray and lying within either plane of projection is imaged point for point by rays lying wholly within that plane of projection. Considering for the moment only the tangential plane of projection, let us choose two such line elements, which we shall call the first and second tangential reference image. Let P_ν and Q_ν respectively be the intercepts which the tangential projection of the ray makes with these two reference images. Their values completely determine the position and orientation of the tangential projection of the ray. Now let M_ν be the magnification with which the νth member images the first tangential reference image, L_ν the corresponding magnification for the second tangential reference image, P_ν' the intercept of the tangential projection of the ray on the first tangential reference image in the image space of the νth member, Q_ν' the corresponding quantity for the second tangential reference image, so that we have:

$$P_\nu' = P_\nu M_\nu, \quad Q_\nu' = Q_\nu L_\nu. \tag{1a}$$

The values of P_ν' and Q_ν' determine the position and orientation of the tangential projection of the ray in the image space of the νth member. Similarly, we choose two reference images in the sagittal plane. Throughout, we shall denote corresponding quantities referring to the sagittal projection by small letters where capitals are used for the tangential projection. And so we have:

$$p_\nu' = p_\nu m_\nu, \quad q_\nu' = q_\nu l_\nu. \tag{1b}$$

While the choice of the four astigmatic reference images is arbitrary in principle, it will be convenient in practice to choose for the first sagittal and tangential reference images two line elements through the corresponding astigmatic images of the point of intersection of the original meridional ray with the plane of the aperture stop, and for the other two, two line elements through the astigmatic images of the axial point of the object plane. With this choice, the distance b_ν of the second sagittal reference image from the axis vanishes.

The reference frame consisting of axis, meridional ray, tangential and sagittal plane, and astigmatic reference images, as shown in Fig. 1, is to be thought of as being rigidly fixed with respect to the member to which it pertains and moving with it. Hence, there are two such reference frames in each of the intermediate spaces of the system: one, pertaining to the image space of the

preceding member, the other, to the object space of the next member. If the latter is perfectly coaxial with its predecessor, the two reference frames coincide.

Now, for the purpose of the present method, it will be convenient to introduce a definition of the measures of decentration and tilt which is rather different from that used by the practical optician. The practical optician takes an arbitrary line in space, such as the axis of a well-made mount for the lenses, to be the intended axis of the system, and he measures all decentrations and tilts with respect to this axis. For our purposes, it will be more convenient to measure the decentration and tilt of each member with respect to its predecessor. Then, decentration and tilt become properties of the space between two members, rather than properties of a member.

Now, if a ray has been traced through the $(\nu-1)$th member, so that $P_{\nu-1}'$, $Q_{\nu-1}'$, $p_{\nu-1}'$, $q_{\nu-1}'$ are known, and if the decentration and tilt of the νth member with respect to the $(\nu-1)$th are known, we must find P_ν, Q_ν, p_ν, q_ν. In specifying decentration and tilt, we may make use of Chasle's theorem: Any infinitesimal displacement of a rigid body can be represented as the result of a translation of an arbitrarily chosen point within the body combined with a rotation about an axis through that point. Only the translation, and not the magnitude or direction of the rotation, depends on the choice of the point. We may therefore specify the tilt in each space about an arbitrarily chosen pivot. Any combination of decentrations and tilts for which we may wish to calculate the aberrations of the system may be represented in terms of decentrations of and tilts about our arbitrarily chosen pivots. For purposes of routine tabulation, we shall make it a practice to choose as the pivot in each space, the apex of the first surface of the member following that space. This has been done in Tables I and II below.

Now let the respective distances of the projections of the first and second tangential reference images and the first and second sagittal reference images upon the axis of the member from the axis of tilt be E_ν, F_ν, e_ν, f_ν (*cf.* Fig. 1). Let α_ν be the angle between the original meridional ray and the axis of the member, β_ν the angle of tilt, g_ν the lateral displacement of the νth member with respect to the $(\nu-1)$th. Let A_ν be the distance of the first tangential reference image from the axis of the member, B_ν the corresponding distance for the second tangential reference image, a_ν and b_ν the corresponding distances for the respective sagittal reference images. Throughout, unprimed quantities refer to the object space of the νth member, primed quantities to the image space of the νth member.

The sign convention for the above quantities will be taken as follows: In the plane of the meridional ray first traced, one direction perpendicular to the axis is arbitrarily taken as up. Then, if the ray is slanting down

TABLE II. Hypergon. Linear and circular pseudo-coma per unit decentration and tilt in each space, for two rays, together with (the lower of each pair of entries) the third-order approximation.*

Radius	+8.47		+8.51	D		−8.51		−8.47	
spacing		2.21		6.80	6.80		2.21		
index		1.517		1	1		1.517		
									Diaphragm
$R_1'=-53.4$									
decentration									
linear		−0.791			+0.0185		+0.816		+0.004
		−0.851			+0.02166		+0.880		+0.005
circular		+0.0388			+0.01062		+0.0258		−0.0002
		+0.0354			+0.01236		−0.0209		−0.0004
tilt									
linear		−19.10			−12.46		−21.05		
		−19.48			−12.35		−21.59		
circular		−13.90			−14.28		−13.91		
		−13.92			−14.29		−13.96		
$R_1'=-135.6$									
decentration									
linear		−3.95			+0.0736		+4.05		−0.0023
		−5.49			+0.1397		+5.68		+0.0342
circular		+0.304			+0.0428		−0.2463		−0.0042
		+0.229			+0.0798		−0.1353		−0.0027
tilt									
linear		−116.0			−82.7		−125.5		
		−125.7			−79.7		−139.4		
circular		−89.3			−92.1		−89.1		
		−89.9			−92.2		−90.1		

* U. S. Patent No. 706650, August 12, 1902, C. P. Goerz.

as it advances, α_ν is positive. A_ν, B_ν, a_ν, and b_ν are positive or negative depending on whether the astigmatic reference images in question lie above or below the axis. P_ν and Q_ν are positive pointing up. p_ν and q_ν are positive in the clockwise direction about the axis of the system as seen by an observer who is looking in the direction in which the light advances. E_ν, F_ν, e_ν, and f_ν are positive when the projections of the corresponding astigmatic reference images are beyond the pivot to which they are referred.

We now introduce right-handed Cartesian coordinates. $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ are unit vectors which are respectively parallel to the x-, y-, z-axis. The latter is parallel to the axis of the undecentered system and positive in the direction in which the light advances. Strictly speaking, each member carries its coordinate system with it as it is decentered; but where there is no ambiguity, the subscripts on the symbols $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ will be omitted. $\mathbf{g}_\nu = g_\nu(\mathbf{i}\cos\varphi_\nu + \mathbf{j}\sin\varphi_\nu)$ is the vector displacement of the apex of the first surface of the νth member with respect to the axis of the $(\nu-1)$th. $\mathbf{v}_\nu = \mathbf{i}\cos\psi_\nu + \mathbf{j}\sin\psi_\nu$ is a unit vector along the axis of tilt, positive in such a direction that for a positive value of β_ν the tilt appears clockwise to an observer looking in the direction of the vector $\mathbf{v}_\nu$. Let $\mathbf{h} = \mathbf{i}\cos\gamma + \mathbf{j}\sin\gamma$ be a unit vector perpendicular to the axis of the system and lying in a plane containing the meridional ray which serves as the axis of our orthogonal system. Finally, let $\mathbf{t}_\nu$ be a unit vector parallel to the tangential axis in the direction of positive P_ν and Q_ν. $\mathbf{k}\times\mathbf{h}$ is a unit vector in the direction of positive p_ν and q_ν. Then we have:

$$\left.\begin{aligned} \mathbf{t}_\nu &= \mathbf{i}\cos\alpha_\nu\cos\gamma + \mathbf{j}\cos\alpha_\nu\sin\gamma + \mathbf{k}\sin\alpha_\nu, \\ \mathbf{k}\times\mathbf{h} &= -\mathbf{i}\sin\gamma + \mathbf{j}\cos\gamma \end{aligned}\right\}. \tag{2}$$

Let $\mathbf{U}_\nu$ be the displacement (due to decentration and tilt combined) of the first tangential reference object to the νth member with respect to the first tangential reference image to the $(\nu-1)$th member; let $\mathbf{u}_\nu$, $\mathbf{W}_\nu$, and $\mathbf{w}_\nu$, respectively, be the corresponding quantities for the first sagittal reference object and the second tangential and sagittal reference objects respectively in the object space of the νth member. Then

$$\left.\begin{aligned} \mathbf{U}_\nu &= \mathbf{g}_\nu + \beta_\nu E_\nu \mathbf{v}_\nu\times\mathbf{k} + \beta_\nu A_\nu \mathbf{v}_\nu\times\mathbf{h} \\ &= \mathbf{i}[g_\nu\cos\varphi_\nu + \beta_\nu E_\nu \sin\psi_\nu] \\ &\quad + \mathbf{j}[g_\nu\sin\varphi_\nu - \beta_\nu E_\nu\cos\psi_\nu] \\ &\quad - \mathbf{k}\beta_\nu A_\nu\sin(\psi_\nu-\gamma) \\ \mathbf{u}_\nu &= \mathbf{g}_\nu + \beta_\nu e_\nu \mathbf{v}_\nu\times\mathbf{k} + \beta_\nu a_\nu \mathbf{v}_\nu\times\mathbf{h} \\ &= \mathbf{i}[g_\nu\cos\varphi_\nu + \beta_\nu e_\nu \sin\psi_\nu] \\ &\quad + \mathbf{j}[g_\nu\sin\varphi_\nu - \beta_\nu e_\nu\cos\psi_\nu] \\ &\quad - \mathbf{k}\beta_\nu a_\nu\sin(\psi_\nu-\gamma) \\ \mathbf{W}_\nu &= \mathbf{g}_\nu + \beta_\nu F_\nu \mathbf{v}_\nu\times\mathbf{k} + \beta_\nu B_\nu \mathbf{v}_\nu\times\mathbf{h} \\ &= \mathbf{i}[g_\nu\cos\varphi_\nu + \beta_\nu F_\nu \sin\psi_\nu] \\ &\quad + \mathbf{j}[g_\nu\sin\varphi_\nu - \beta_\nu F_\nu\cos\psi_\nu] \\ &\quad - \mathbf{k}\beta_\nu B_\nu\sin(\psi_\nu-\gamma) \\ \mathbf{w}_\nu &= \mathbf{g}_\nu + \beta_\nu f_\nu \mathbf{v}_\nu\times\mathbf{k} + \beta_\nu b_\nu \mathbf{v}_\nu\times\mathbf{h} \\ &= \mathbf{i}[g_\nu\cos\varphi_\nu + \beta_\nu f_\nu \sin\psi_\nu] \\ &\quad + \mathbf{j}[g_\nu\sin\varphi_\nu - \beta_\nu f_\nu\cos\psi_\nu] \\ &\quad - \mathbf{k}\beta_\nu b_\nu\sin(\psi_\nu-\gamma) \end{aligned}\right\}, \tag{3}$$

$$\left.\begin{aligned} P_{\nu-1}' - P_\nu &= \mathbf{U}_\nu\cdot\mathbf{t}_\nu = g_\nu\cos\alpha_\nu\cos(\varphi_\nu-\gamma) \\ &\quad + \beta_\nu E_\nu\cos\alpha_\nu\sin(\psi_\nu-\gamma) \\ &\quad - \beta_\nu A_\nu\sin\alpha_\nu\sin(\psi_\nu-\gamma) \\ p_{\nu-1}' - p_\nu &= \mathbf{u}_\nu\cdot\mathbf{k}\times\mathbf{h} = g_\nu\sin(\varphi_\nu-\gamma) \\ &\quad - \beta_\nu e_\nu\cos(\psi_\nu-\gamma) \\ Q_{\nu-1}' - Q_\nu &= \mathbf{W}_\nu\cdot\mathbf{t} = g_\nu\cos\alpha_\nu\cos(\varphi_\nu-\gamma) \\ &\quad + \beta_\nu F_\nu\cos\alpha_\nu\sin(\psi_\nu-\gamma) \\ &\quad - \beta_\nu B_\nu\sin\alpha_\nu\sin(\psi_\nu-\gamma) \\ q_{\nu-1}' - q_\nu &= \mathbf{w}_\nu\cdot\mathbf{k}\times\mathbf{h} = g_\nu\sin(\varphi_\nu-\gamma) \\ &\quad - \beta_\nu f_\nu\cos(\psi_\nu-\gamma) \end{aligned}\right\}. \tag{4}$$

Combining Eqs. (1a), (1b), and (4), and summing over all spaces, we have:

$$\left.\begin{aligned} P_\mu' &= -M_{\zeta,\mu+1}[\bar g_\zeta\cos\alpha_\zeta\cos(\bar\varphi_\zeta-\gamma) \\ &\quad + \bar\beta_\zeta E_\zeta\cos\alpha_\zeta\sin(\bar\psi_\zeta-\gamma) \\ &\quad - \bar\beta_\zeta A_\zeta\sin\alpha_\zeta\sin(\bar\psi_\zeta-\gamma)] \\ &\quad - \sum_{\nu=\zeta+1}^{\mu} M_{\nu,\mu+1}[g_\nu\cos\alpha_\nu\cos(\varphi_\nu-\gamma) \\ &\quad + \beta_\nu E_\nu\cos\alpha_\nu\sin(\psi_\nu-\gamma) \\ &\quad - \beta_\nu A_\nu\sin\alpha_\nu\sin(\psi_\nu-\gamma)] \\ p_\mu' &= -m_{\zeta,\mu+1}[\bar g_\zeta\sin(\bar\varphi_\zeta-\gamma) - \bar\beta_\zeta e_\zeta\cos(\bar\psi_\zeta-\gamma)] \\ &\quad - \sum_{\nu=\zeta+1}^{\mu} m_{\nu,\mu+1}[g_\nu\sin(\varphi_\nu-\gamma) \\ &\quad - \beta_\nu e_\nu\cos(\psi_\nu-\gamma)] \\ Q_\mu' &= -L_{1,\mu+1}g_1\cos\alpha_1\cos(\varphi_1-\gamma) \\ &\quad - \sum_{\nu=2}^{\mu} L_{\nu,\mu+1}[g_\nu\cos\alpha_\nu\cos(\varphi_\nu-\gamma) \\ &\quad + \beta_\nu F_\nu\cos\alpha_\nu\sin(\psi_\nu-\gamma) \\ &\quad - \beta_\nu B_\nu\sin\alpha_\nu\sin(\psi_\nu-\gamma)] \\ q_\mu' &= -l_{1,\mu+1}g_1\sin(\varphi_1-\gamma) \\ &\quad - \sum_{\nu=2}^{\mu} l_{\nu,\mu+1}[g_\nu\sin(\varphi_\nu-\gamma) \\ &\quad - \beta_\nu f_\nu\cos(\psi_\nu-\gamma)] \end{aligned}\right\}. \tag{5}$$

The μth member is the last one of the system. For the sake of uniformity of notation, we take into account the possibility of a decentration of the object itself by admitting a decentration $\mathbf{g}_1 = g_1(\mathbf{i}\cos\varphi_1 + \mathbf{j}\sin\varphi_1)$ of the first member with respect to the original position of the axis of the undecentered system in the object space. A tilt of the object plane about its axial point need not be considered, because it will only give rise to infinitesimal terms of higher order. A tilt in the object space of the system as a whole and about the apex of the first surface of the system can therefore always be represented as an equivalent decentration.

Similarly, we are admitting the possibility of a decentration and tilt of the aperture stop, which we suppose located in the object space of the ζth member, and we take this into account by denoting the decentration of the ζth member with respect to the aperture stop by $\bar{\mathbf{g}}_\zeta = \bar g_\zeta(\mathbf{i}\cos\bar\varphi_\zeta + \mathbf{j}\sin\bar\varphi_\zeta)$ and by measuring the tilt $\bar\beta_\zeta$

of the ζth member with respect to the aperture stop about an axis of azimuth $\bar{\psi}_\zeta$ through the apex of the first surface of the ζth member. g_ζ, φ_ζ, β_ζ, and ψ_ζ retain their original significance. In practice, where we will usually be given the decentration and tilt of aperture stop or object plane, these are simply the negatives of the respective quantities used here. Incidentally, we have $E_\zeta = e_\zeta$, by definition.

It will be noted that the decentrations and tilts in the spaces preceding the aperture stop do not contribute to P_μ' and p_μ'. The reason for this is that we have specified the position of the point of intersection of the ray with the plane of the aperture stop independently of the decentrations and tilts. In this manner, we specify the choice of the ray to be traced. This is important: It should be a foregone conclusion that the outcome of a ray trace depends on the choice of the ray to be traced. It is understood throughout that the system is free from vignetting. Otherwise, some of the rays traced by this method may be cut off.

Finally, it should be pointed out that in Eqs. (5) we have used the simplified notation:

$$\left.\begin{aligned} M_{\nu,\mu+1} &= M_\nu M_{\nu+1} M_{\nu+2} \cdots M_{\mu-1} M_\mu \\ m_{\nu,\mu+1} &= m_\nu m_{\nu+1} m_{\nu+2} \cdots m_{\mu-1} m_\mu \\ L_{\nu,\mu+1} &= L_\nu L_{\nu+1} L_{\nu+2} \cdots L_{\mu-1} L_\mu \\ l_{\nu,\mu+1} &= l_\nu l_{\nu+1} l_{\nu+2} \cdots l_{\mu-1} l_\mu \end{aligned}\right\}. \tag{6}$$

We are now in a position to write down an expression for the vector $\mathbf{R}_\mu'$ from the axial point of the paraxial image plane to the point of intersection of the ray with this plane. Let G be the distance from the apex of the last surface of the last member to this plane, positive if the light reaches the last surface first (i.e., if the image is real). In view of our definition of decentration and tilt, the image plane is of course perpendicular to the axis of the last member and remains in the same position with respect to the last member, regardless of the decentrations and tilts introduced. For the sake of uniformity of notation, we shall denote by E_μ', F_μ', e_μ', f_μ' the distances of the projections of the respective astigmatic reference images upon the axis of the last (μth) member beyond the last surface of the last member. Then we have:

$$\begin{aligned} \mathbf{R}_\mu' = \mathbf{h}(f_\mu' - G)\tan\alpha_{\mu+1} &+ [P_\mu'(F_\mu' - G)/(F_\mu' - E_\mu') \\ &+ Q_\mu'(G - E_\mu')/(F_\mu' - E_\mu')]\mathbf{h}/\cos\alpha_{\mu+1} \\ &+ [p_\mu'(f_\mu' - G)/(f_\mu' - e_\mu') \\ &+ q_\mu'(G - e_\mu')/(f_\mu' - e_\mu')]\mathbf{k}\times\mathbf{h}. \end{aligned} \tag{7}$$

Substituting from Eqs. (5) and gathering together the terms in the individual decentrations and tilts, we have:

$$\begin{aligned} \mathbf{R}_\mu' = \mathbf{h}(f_\mu' - G)\tan\alpha_{\mu+1} &+ \sum\{[\Theta_\nu g_\nu \cos(\varphi_\nu - \gamma) + \Xi_\nu \beta_\nu \sin(\psi_\nu - \gamma)]\mathbf{h} \\ &+ [\theta_\nu g_\nu \sin(\varphi_\nu - \gamma) - \xi_\nu \beta_\nu \cos(\psi_\nu - \gamma)]\mathbf{k}\times\mathbf{h}\}, \end{aligned} \tag{8}$$

where the summation sign is to be interpreted broadly to include the terms in g_1, $\bar{g}_\zeta$, and $\bar{\beta}_\zeta$. For the coefficients, we have:

$$\left.\begin{aligned} \Xi_1 &= \xi_1 = 0 \\ \Theta_1 &= -L_{1,\mu+1}(G - E_\mu')\cos\alpha_1/(F_\mu' - E_\mu')\cos\alpha_{\mu+1} \\ \theta_1 &= -l_{1,\mu+1}(G - e_\mu')/(f_\mu' - e_\mu') \end{aligned}\right\}. \tag{9a}$$

When $1 < \nu < \zeta$:

$$\left.\begin{aligned} \Theta_\nu &= -L_{\nu,\mu+1}(G - E_\mu')\cos\alpha_\nu/(F_\mu' - E_\mu')\cos\alpha_{\mu+1} \\ \theta_\nu &= -l_{\nu,\mu+1}(G - e_\mu')/(f_\mu' - e_\mu') \\ \Xi_\nu &= -L_{\nu,\mu+1}(F_\nu\cos\alpha_\nu - B_\nu\sin\alpha_\nu) \\ &\qquad \times (G - E_\mu')/(F_\mu' - E_\mu')\cos\alpha_{\mu+1} \\ \xi_\nu &= -l_{\nu,\mu+1} f_\nu (G - e_\mu')/(f_\mu' - e_\mu') \end{aligned}\right\}. \tag{9b}$$

When $\nu = \zeta$, we have two kinds of terms: those in g_ζ and β_ζ, whose coefficients Θ_ζ, θ_ζ, Ξ_ζ, ξ_ζ are defined as for $1 < \nu < \zeta$, and those in $\bar{g}_\zeta$ and $\bar{\beta}_\zeta$, whose coefficients $\bar{\Theta}_\zeta$, $\bar{\theta}_\zeta$, $\bar{\Xi}_\zeta$, $\bar{\xi}_\zeta$ are given by the formulas:

$$\left.\begin{aligned} \bar{\Theta}_\zeta &= -M_{\zeta,\mu+1}(F_\mu' - G)\cos\alpha_\zeta/(F_\mu' - E_\mu')\cos\alpha_{\mu+1} \\ \bar{\theta}_\zeta &= -m_{\zeta,\mu+1}(f_\mu' - G)/(f_\mu' - e_\mu') \\ \bar{\Xi}_\zeta &= -M_{\zeta,\mu+1}(E_\zeta\cos\alpha_\zeta - A_\zeta\sin\alpha_\zeta) \\ &\qquad \times (F_\mu' - G)/(F_\mu' - E_\mu')\cos\alpha_{\mu+1} \\ \bar{\xi}_\zeta &= -m_{\zeta,\mu+1} e_\zeta (f_\mu' - G)/(f_\mu' - e_\mu') \end{aligned}\right\}. \tag{9c}$$

Finally, when $\nu > \zeta$, we have:

$$\left.\begin{aligned} \Theta_\nu &= [-M_{\nu,\mu+1}(F_\mu' - G) - L_{\nu,\mu+1}(G - E_\mu')] \\ &\qquad \times\cos\alpha_\nu/(F_\mu' - E_\mu')\cos\alpha_{\mu+1} \\ \theta_\nu &= [-m_{\nu,\mu+1}(f_\mu' - G) \\ &\qquad - l_{\nu,\mu+1}(G - e_\mu')]/(f_\mu' - e_\mu') \\ \Xi_\nu &= [-M_{\nu,\mu+1}(E_\nu\cos\alpha_\nu - A_\nu\sin\alpha_\nu)(F_\mu' - G) \\ &\qquad - L_{\nu,\mu+1}(F_\nu\cos\alpha_\nu - B_\nu\sin\alpha_\nu) \\ &\qquad \times (G - E_\mu')]/(F_\mu' - E_\mu')\cos\alpha_{\mu+1} \\ \xi_\nu &= [-m_{\nu,\mu+1} e_\nu (f_\mu' - G) \\ &\qquad - l_{\nu,\mu+1} f_\nu (G - e_\mu')]/(f_\mu' - e_\mu') \end{aligned}\right\}. \tag{9d}$$

Finally, substituting the expressions for $\mathbf{h}$ and $\mathbf{k}\times\mathbf{h}$ into (8), we have:

$$\begin{aligned} \mathbf{R}_\mu' = \mathbf{h}(f_\mu' - G)\tan\alpha_{\mu+1} &+ \tfrac{1}{2}\mathbf{i}\sum\{g_\nu[(\Theta_\nu + \theta_\nu)\cos\varphi_\nu \\ &+ (\Theta_\nu - \theta_\nu)\cos(\varphi_\nu - 2\gamma)] \\ &+ \beta_\nu[(\Xi_\nu + \xi_\nu)\sin\psi_\nu + (\Xi_\nu - \xi_\nu)\sin(\psi_\nu - 2\gamma)]\} \\ &+ \tfrac{1}{2}\mathbf{j}\sum\{g_\nu[(\Theta_\nu + \theta_\nu)\sin\varphi_\nu \\ &- (\Theta_\nu - \theta_\nu)\sin(\varphi_\nu - 2\gamma)] \\ &+ \beta_\nu[-(\Xi_\nu + \xi_\nu)\cos\psi_\nu \\ &\qquad + (\Xi_\nu - \xi_\nu)\cos(\psi_\nu - 2\gamma)]\}. \end{aligned} \tag{10}$$

It turns out that the contributions which the individual decentrations and tilts make to the final displacement are simply additive as vectors and can be studied individually. The first term in the right-hand member of Eq. (10) is the spherical aberration of the system. This is independent of the decentrations and tilts and is of no further interest. In any usable optical instrument, it is always small compared to the field of view. For this reason, we shall not discuss the possibility of a tilt of the image plane. The terms in $\frac{1}{2}(\Theta_\nu + \theta_\nu)$ together represent a vector which is in the nature of a lateral displacement of the image. While such a displacement would also have taken place within the paraxial domain, the actual

displacement will exceed the paraxial displacement by an amount (positive or negative) which we shall call linear coma.

The terms in $\frac{1}{2}(\Theta_\nu-\theta_\nu)$ together represent a vector whose magnitude is independent of φ_ν or γ, but whose azimuth is given by $\varphi_\nu-2\gamma$. Hence, other things being equal, as the point of intersection of the ray with the plane of the aperture stop describes a circle about the center of the aperture stop, the vector in question traces out a circle twice. This is a well known characteristic of coma. The quantity $\frac{1}{2}(\Theta_\nu-\theta_\nu)$ will be referred to as circular coma. It can be shown from Maréchal's formulas[6] that, within the third-order domain, linear coma equals twice the circular coma.

The terms in $\frac{1}{2}(\Xi_\nu+\xi_\nu)$ and $\frac{1}{2}(\Xi_\nu-\xi_\nu)$, respectively, can be interpreted similarly to those in $\frac{1}{2}(\Theta_\nu+\theta_\nu)$ and $\frac{1}{2}(\Theta_\nu-\theta_\nu)$.

As a numerical example, we show in Table I the data of construction of a Tessar-type photographic objective, together with the linear and circular coma per unit decentration and tilt in each space. The design has been ratioed to give a focal length of 100. The calculations have been performed by the above method for two rays, a marginal ray, $K=11$, and a zonal ray, $K=8$, where K is the distance from the axis in the object space at which the ray intersects the entrance pupil plane. Under each entry, we have given the third-order approximation to that entry. These latter quantities have been calculated by Maréchal's method.[6] Note that Maréchal's formulas give the wave-front aberration. The transverse aberration can be obtained from this with the aid of the rule that the ray is perpendicular to the wave front.

The quantities $\frac{1}{2}(\overline{\Theta}_\xi+\bar{\theta}_\xi)$, $\frac{1}{2}(\overline{\Theta}_\xi-\bar{\theta}_\xi)$, $\frac{1}{2}(\overline{\Xi}_\xi+\bar{\xi}_\xi)$, $\frac{1}{2}(\overline{\Xi}_\xi-\bar{\xi}_\xi)$ are shown in the last column. These are always small for a well-corrected system. The aberrations for decentration of the object have been omitted, since they present nothing new.

A table such as this enables us to answer the practical optician's question, to what extent the effect of decentration or tilt in one space may be offset with the effect of decentration or tilt in another space. Within the third-order domain, that is always possible. And it turns out that, by and large, the third-order aberrations give a close approximation for the present system. There are a few exceptions.

WIDE-FIELD SYSTEMS

The method used for calculating the aberrations of a slightly decentered wide-aperture system with a narrow field of view is easily adapted to the problem of calculating the aberrations of a slightly decentered wide-field system with a narrow aperture. As before, we start out by supposing the system to be perfectly coaxial, and, in addition, we shall consider the object and image planes to be perpendicular to the axis of the system. A ray is traced from an object point at a finite distance from the axis through the center of the aperture stop. Sagittal and tangential rays near this first ray are traced, first through the object point, then through the center of the aperture stop, and, as before, the optical properties of the orthogonal system about the meridional ray are calculated for each member of the system. The quantities $A_\nu, B_\nu, E_\nu, e_\nu, F_\nu, f_\nu, P_\nu, p_\nu, Q_\nu, q_\nu, L_\nu, l_\nu, M_\nu, m_\nu$ have the same meaning as before. We no longer need to consider the possibility of a decentration of the object plane, since this would be tantamount to pointing the system at a different object. But, a tilt of the object or image plane, showing the effect of imperfect "squaring-on," now becomes important. The possibility of a decentration of the aperture stop will be taken into account as before. A tilt of the infinitesimal aperture stop about its center, through an infinitesimal angle, will give rise to only higher order infinitesimals and may be neglected. All this leads to the following modification of Eqs. (5):

$$\left.\begin{aligned}
P_\mu' &= -M_{\xi,\mu+1}\vartheta_\xi\cos\alpha_\xi\cos(\bar{\varphi}_\xi-\gamma)\\
&\quad-\sum_{\nu=\xi+1}^{\mu+1}M_{\nu,\mu+1}[g_\nu\cos\alpha_\nu\cos(\varphi_\nu-\gamma)\\
&\qquad+\beta_\nu E_\nu\cos\alpha_\nu\sin(\psi_\nu-\gamma)\\
&\qquad-\beta_\nu A_\nu\sin\alpha_\nu\sin(\psi_\nu-\gamma)]\\
p_\mu' &= -m_{\xi,\mu+1}\vartheta_\xi\sin(\bar{\varphi}_\xi-\gamma)\\
&\quad-\sum_{\nu=\xi+1}^{\mu+1}m_{\nu,\mu+1}[g_\nu\sin(\varphi_\nu-\gamma)\\
&\qquad-\beta_\nu e_\nu\cos(\psi_\nu-\gamma)]\\
Q_\mu' &= -\sum_{\nu=1}^{\mu+1}L_{\nu,\mu+1}[g_\nu\cos\alpha_\nu\cos(\varphi_\nu-\gamma)\\
&\qquad+\beta_\nu F_\nu\cos\alpha_\nu\sin(\psi_\nu-\gamma)\\
&\qquad-\beta_\nu B_\nu\sin\alpha_\nu\sin(\psi_\nu-\gamma)]\\
q_\mu' &= -\sum_{\nu=1}^{\mu+1}l_{\nu,\mu+1}[g_\nu\sin(\varphi_\nu-\gamma)\\
&\qquad-\beta_\nu f_\nu\cos(\psi_\nu-\gamma)]
\end{aligned}\right\}. \tag{5a}$$

We have taken some liberties with the notation in these equations: the terms for which $\nu=1$ pertain to a decentration and tilt of the first member with respect to a coordinate system fixed in the object plane. The axis of tilt passes through the apex of the first surface of the system. Similarly, the terms for $\nu=\mu+1$ pertain to a decentration and tilt of a coordinate system fixed in the image plane, with respect to the last member. Since there is no $(\mu+1)$th member, we make an exception to our usual rule by taking the axis of tilt through the apex of the last surface of the system. Since we are only concerned with a pure tilt about the axial point in the case of object- and image-plane, we have the further conditions:

$$\left.\begin{aligned}
F_1\beta_1\mathbf{v}_1\times\mathbf{k} &= -\mathbf{g}_1\\
G\beta_{\mu+1}\mathbf{v}_{\mu+1}\times\mathbf{k} &= -\mathbf{g}_{\mu+1}
\end{aligned}\right\}. \tag{5b}$$

Note that

$$F_1=f_1,$$
$$M_{\mu+1,\mu+1}=m_{\mu+1,\mu+1}=L_{\mu+1,\mu+1}=l_{\mu+1,\mu+1}=1.$$

Equations (7), (8), (9b), and (9d), remain valid without

change. In place of (9a) we use (9b). Equation (9d) also applies for $\nu=\mu+1$. In these last two cases, we must express g_ν in terms of β_ν with the aid of Eqs. (5b). The first two of Eqs. (9c) remain unchanged. The other two are not needed, since $\bar{\beta}_t=0$. Equation (10) remains unchanged in form. But, while γ, the azimuth of the original meridional ray, was also the azimuth of the point of intersection of the ray with the plane of the aperture stop in the case of wide-aperture systems, it is now the azimuth of the object point. The form of the aberration has not changed, but its geometrical significance has changed from coma to an asymmetrical distortion. Because of this similarity, we propose to call this aberration pseudo-coma, and we shall distinguish between linear and circular pseudo-coma by analogy with the linear and circular coma defined above. This aberration was first described by Conrady[1] within the third-order domain. It is known to photo-grammetrists as tangential distortion.

Again we obtain formulas for the third-order aberrations from those of Maréchal, as outlined above. Note that the term which represents a tilt of the image plane (Conrady calls this the error of squaring-on) can also be represented as the resultant of linear and circular pseudo-coma, the two being equal.

If we wish to compare the third-order aberrations with the actual aberrations, it is important that we choose the proper independent variables, since the higher order aberrations will depend on this choice. The reason for this is that we express the third-order aberrations in terms of paraxial quantities. Within the paraxial region, some of these quantities may be proportional to other quantities, while such proportionality may no longer exist outside the paraxial domain. If such proportionality is nonetheless assumed and a change of independent variable is made accordingly, the values of the third-order aberrations will change for large values of the independent variable.

The choice of $\mathbf{R}_\mu'$ as an independent variable suggests itself as reasonable, because then, if linear and circular pseudo-coma are equal (concerning the directions of these vectors, see below) and proportional to $R_\mu'^2$, the resulting aberration is a pure error of squaring-on, no matter how large R_μ' may be. The proof is left to the reader. This aberration will transform a straight line into a straight line, but it transforms parallel lines into converging lines, unless the original lines are parallel to the vector $\mathbf{v}_\mu'$.

It follows from this definition that a tilt only of the plate holder of a camera does not necessarily cause a pure error of squaring-on. Other aberrations will be present if the exit pupil is imaged with spherical aberration. However, in practice, this error will usually be minute.

Again, the question arises whether the effect of one decentration or tilt may be offset with the effect of another. The answer is not as simple as in the case of wide-aperture systems, because there is no simple relationship between linear and circular pseudo-coma, even within the third-order domain. If only a single decentration or tilt is present, then, for $\gamma=0$, the azimuth of the circular pseudo-coma is minus the azimuth of the linear pseudo-coma (*cf.* Eq. (10)). If γ equals the azimuth of the linear pseudo-coma, it also equals the azimuth of the circular pseudo-coma. But if several decentrations and tilts of random magnitude and direction are present, the magnitudes and directions of linear and circular pseudo-coma are unpredictable. If the maker of the optical system were in a position to control just one decentration, or one tilt, he would not have enough independent variables at his disposal to reduce both linear and circular pseudo-coma to zero. However, if the magnitudes of the contributions to linear and circular pseudo-coma which he can control are not equal (the relationship between their azimuths being as just explained for a single decentration or tilt), he can render the resultants equal in magnitude and interrelated in azimuth as if they were due to a single decentration. In other words, if the decentration or tilt which can be controlled gives rise to an asymmetrical distortion other than a pure error of squaring-on, it can be used to reduce any given pseudo-coma to a pure error of squaring-on. The latter may be offset with a properly controlled tilt of the image plane (plate holder of the camera). Of course, these conditions can be satisfied only for either the third-order domain or a single zone of the image plane.

Table II gives the data of construction, and aberrations due to decentration, for a hypergon, ratioed to give a focal length of 100. The linear and circular pseudo-coma and the corresponding third-order aberrations per unit decentration and tilt in each space are given for two principal rays, inclined respectively at 27°.9 and 53°.4 to the axis of the original (undecentered) system. The conventions used in this table are in every way analogous to those used in Table I.

The distinction between decentration and tilt is not necessary unless the system contains aspheric surfaces. Otherwise, we can express one in terms of the other, and in this manner we may check our results for errors: If two successive spaces are separated by a spherical surface, and both are tilted through the same angle, but in opposite directions, about an axis through the center of curvature of the surface in question, the resulting aberrations should be zero.

The author wishes to thank Mr. G. A. Smythe for help with the computations.

Note added in proof.—The error of squaring-on also arises in every one of the intermediate spaces, due to tilt of the object plane of the νth member with respect to the image plane of the $(\nu-1)$th. It must be taken into account as part and parcel of the third-order pseudo-coma if the comparison with the result obtained by our method is to be valid. Maréchal (see reference 6) did not mention it because he considered it a purely geometrical rather than optical effect. He will treat it, however, in a forthcoming paper in Revue d'Optique. It was also taken into account in Eq. (10) of the paper by W. G. Stephan quoted above.

Some Properties of Thin Evaporated Films on Glass*

JOSEPH L. ROOD**
Department of Physics, University of California, Berkeley, California
(Received April 21, 1949)

Properties of thin evaporated films of MgF_2 and of $CaSiO_3$ on glass were studied. Thicknesses of the films were determined from measurements of reflected intensity as a function of wave-length, and indices of refraction computed. Density measurements were carried out for MgF_2 films. All these properties were found to be different from those of the corresponding bulk materials, and furthermore, the properties of the films change with time. Results are given for indices of refraction, densities, and aging effects.

I. INTRODUCTION

THIN evaporated films on glass (films whose optical thicknesses are of the order of a quarter- or half-wave-length of visible light) have been studied by various observers for some years. In general, the properties of a substance in the form of a thin film on glass are different from those of the substance in bulk; in the work herein described certain quantitative results were obtained for the two particular films MgF_2 and $CaSiO_3$ which illustrate the differences. Some of the research was concerned with reflection-reducing coatings using a combination of the two films.

II. APPARATUS

The films were prepared by evaporating in vacuum a known amount of the bulk substance from a V-shaped molybdenum heater. A small piece of ordinary crown glass was suspended horizontally ten inches above the bottom of the heater. This arrangement was found to give a deposit on the glass that was very nearly uniform.

The vacuum chamber was a large cylindrical tank, evacuated by a mechanical pump in series with two oil diffusion pumps. The glass was carefully cleaned before being placed in the chamber; final cleaning occurred from a high tension discharge, during evacuation. Evaporations were carried out at pressures of 0.1 to 0.2 micron.

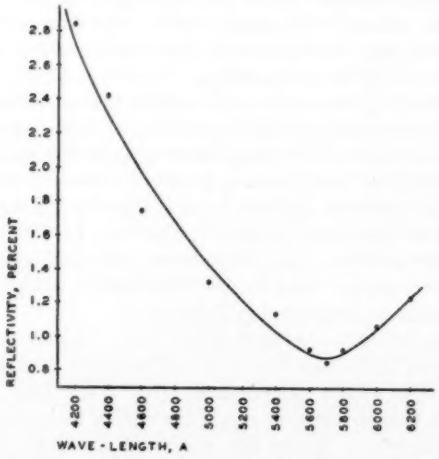

FIG. 1. Experimental values of reflectivity as a function of wave-length for a typical MgF_2 film, M14.

The thickness of the films was found to vary approximately linearly with the amount of substance evaporated. Thus the desired thickness could be produced by placing a certain mass of the substance to be evaporated in the heater, and evaporating the whole amount. Measurements of the thicknesses were obtained from the positions of the maxima or minima of curves of reflected intensity (reflectivity) as a function of wave-length. Reflectivities were measured by comparing readings with an RCA 931A photo-multiplier tube (linear response in region used) for a plain glass with those for a coated glass. Reflection from the rear surfaces of the glasses was eliminated by blackening with a heavy lacquer. The region of the spectrum investigated was 4200A to 6200A; a narrow band was used, of width about 25A near the former wave-length, and near the latter, about 200A.

The reflection-measuring instrument used an angle of incidence of 15°. Normal incidence was assumed, however, since calculations showed that the errors introduced by such an assumption do not exceed one

TABLE I. Indices of refraction and thicknesses of MgF_2 films having quarter-wave minima of reflectivity in region 4200A to 6200A.

1 Number	2 Mass evap., mg	3 Position min. R, A (exp.)	4 Value min. R, % (exp.)	5 Index of refraction (calc.)	6 Physical thickness, A (calc.)
M1	34	4700	0.66	1.33_9	877
M2	35	5000	0.65	1.33_7	932
M3	37	5200	0.72	1.34_3	968
M4	38	5500	1.00	1.36_3	1009
M5	38	5400	0.98	1.32_2	991
M6	38	5800	0.80	1.34_9	1075
M7	38	5600	0.77	1.34_6	1040
M8	38	5400	0.93	1.35_3	994
M9	38	5700	0.99	1.36_2	1046
M10	38	5400	0.64	1.33_6	1011
M11	38	5500	1.22	1.37_7	998
M12	40	5500	0.92	1.35_7	1013
M13	40	6000	0.78	1.34_8	1115
M14	40	5700	0.86	1.35_2	1054
M15	40	5400	0.65	1.33_7	1010

* The report given here is a condensation of a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of California, 1948.

** Now at the University of San Francisco, San Francisco, California.

percent. The index of refraction and its variation with wave-length was known for the uncoated glass, so that reflectivities for coated glasses could be calculated from the currents found experimentally for standard and unknown.

III. OPTICAL PROPERTIES OF THE FILMS

Theory

A brief resume of certain theoretical results will now be given. These results have been derived in several recent publications on the subject.[1] We are interested only in the variation of reflectivities with wave-length, film thicknesses, and film indices of refraction, for light normally incident on glass coated with one or two films. In what follows, the subscript zero refers to the glass, one to the first film (nearest the glass), and two the second film (on top). The ambient medium is assumed to be air of index of refraction one.

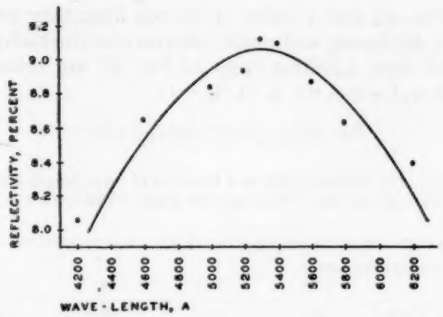

FIG. 2. Experimental values of reflectivity as a function of wave-length for a typical $CaSiO_3$ film, C9.

The reflectivity, R, for one film on glass is given by

$$R=\frac{(n_0+n_1)^2(n_1-1)^2+(n_1+1)^2(n_0-n_1)^2+2(n_1^2-1)(n_0^2-n_1^2)\cos x}{(n_0+n_1)^2(n_1+1)^2+(n_1-1)^2(n_0-n_1)^2+2(n_1^2-1)(n_0^2-n_1^2)\cos x}, \tag{1}$$

where n_0 is the index of the glass, n_1 that of the film, and x_1 is $(2\pi/\lambda)(2n_1d_1)$, i.e., the difference in phase of successively reflected rays from the film of thickness d_1. This equation has two interesting special forms: If the optical thickness n_1d_1 is $k\lambda/4$, where k is an *even* integer,

$$R=[(n_0-1)/(n_0+1)]^2. \tag{2}$$

Thus, the reflectivity of glass with a half-wave film on it is the same as for glass alone, regardless of the index of the film. If k is an *odd* integer, so that $n_1d_1=\lambda/4$, $3\lambda/4$, $\cdots$, however,

$$R=(n_1^2-n_0)^2/(n_1^2+n_0)^2. \tag{3}$$

For $n_1>n_0$, as with $CaSiO_3$, this gives a maximum of reflected intensity; for $n_1<n_0$, as with MgF_2, a minimum. In particular, when the optical thickness is quarter-wave and $n_1^2=n_0$, R is zero. When Eq. (3) is solved for n_1 it takes the form

$$n_1=[n_0(1+(R)^{\frac{1}{2}})/(1-(R)^{\frac{1}{2}})]^{\frac{1}{2}}. \tag{4}$$

For two films on glass, the reflectivity for normal incidence is

$$R=\frac{\left[n_1n_2(1-n_0)-(n_1^2-n_0n_2^2)\tan\frac{x_1}{2}\tan\frac{x_2}{2}\right]^2}{\left[n_1n_2(1+n_0)-(n_1^2+n_0n_2^2)\tan\frac{x_1}{2}\tan\frac{x_2}{2}\right]^2}+\frac{\left[n_2(n_0-n_1^2)\tan\frac{x_1}{2}+n_1(n_0-n_2^2)\tan\frac{x_2}{2}\right]^2}{\left[n_2(n_0+n_1^2)\tan\frac{x_1}{2}+n_1(n_0+n_2^2)\tan\frac{x_2}{2}\right]^2}. \tag{5}$$

TABLE II. Indices of refraction of $CaSiO_3$ films having quarter-wave maxima of reflectivity in region 4200A to 6200A.

1 Number	2 Mass evap., mg	3 Position max. R, A (exp.)	4 Value max. R, % (exp.)	5 Index of refraction (calc.)
C1	22	4300	9.7	1.70_5
C2	24	4500	9.8	1.70_8
C3	24	5000	9.4	1.69_4
C4	25	5600	9.4	1.69_2
C5	25	5700	9.1	1.68_2
C6	26	5100	9.4	1.69_4
C7	28	6000	9.2	1.68_4
C8	28	5200	9.3	1.69_0
C9	28	5300	9.1	1.68_5
C10	28	6000	9.0	1.67_8

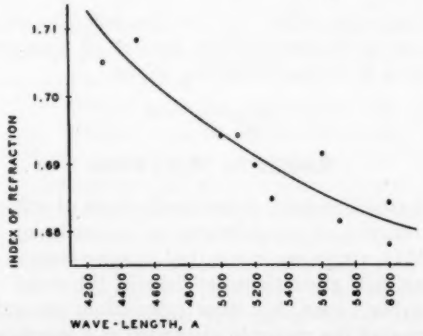

FIG. 3. Indices of refraction of $CaSiO_3$ films having quarter-wave maxima of reflectivity in the region 4200A to 6200A. (A plot of columns 3 and 5 of Table II.)

[1] See, for example, Doris L. Cabellero, J. Opt. Soc. Am. **37**, 176 (1947); P. King and L. B. Lockhart, J. Opt. Soc. Am. **36**, 513 (1946); R. L. Mooney, J. Opt. Soc. Am. **35**, 574 (1945); and Antonin Vašiček, J. Opt. Soc. Am. **37**, 623–634 (1947).

Here n_2, d_2, and x_2 refer to the top film: they are its index, thickness, and phase difference of successive reflected rays. Limiting cases of Eq. (5) are those for which $n_1d_1=n_2d_2=\lambda/4, 3\lambda/4, \cdots$:

$$R=(n_2^2n_0-n_1^2)^2/(n_2^2n_0+n_1^2)^2; \quad (6)$$

TABLE III. Reflectivities as a function of wave-length for two-film coatings on glass. Film nearest glass: $CaSiO_3$, film on top: MgF_2.

Amounts evaporated:	CM1	CM2	CM3	CM4
mg $CaSiO_3$	26	26	26	24
mg MgF_2	40	38	35	35
Reflectivities in percent for various wave-lengths:				
4200A	7.0_8	5.7_8	1.9_5	0.9_1
4400A	5.2_8			
4600A	4.1_5	2.7_2	1.2_7	0.2_0
4800A				0.1_7
5000A	2.2_4	1.8_2	0.6_0	0.0_7
5200A			0.6_0	0.0_7
5400A	1.1_6	0.7_0	0.1_8	0.2_0
5600A	0.9_4	0.2_9	0.1_6	
5800A	0.6_8	0.3_0	0.2_1	0.3_3
6000A	0.6_1			
6200A	1.4_4	1.0_1	1.0_3	0.6_2

TABLE IV. Density of MgF_2 films.

1 Number	2 Mass evap., mg	3 Weight film, mg	4 Thickness, A[a]	5 Area cm^2	6 Density g/cm^3
M5	38	1.12	991	48.0	2.3_6
M12	40	1.33	1013	49.7	2.6_4
M16	50	1.63	1332	50.0	2.4_5
M17	70	2.61	2036	49.9	2.5_7
M18	75	2.94	2110	49.9	2.7_9
M19	82	2.76	2146	50.4	2.5_8
Mean density: (2.56 ± 0.04) g/cm³					

[a] The last four entries of column 4 have been calculated from the "average" index of refraction, 1.35_1, and experimental values of the optical thickness.

or for which $n_1d_1=n_2d_2=\lambda/2, \lambda, \cdots$:

$$R=(n_0-1)^2/(n_0+1)^2. \quad (7)$$

Once again for half-wave films the reflecting power is the same as if there were no films at all. One possible zero form is obtained from Eq. (6) as

$$n_2^2n_0-n_1^2=0. \quad (8)$$

Results for MgF_2 Films

The experimentally determined values of reflectivity as a function of wave-length for a particular MgF_2 film, M14, are shown in Fig. 1. The wave-lengths of the incident light given there are those of the center of the narrow band used. For these calculations the index of refraction of the uncoated glass used as a standard was taken as a constant, 1.52_0; this gives a reflectivity for normal incidence of 4.26 percent. For the film M14 there is a minimum of reflectivity near 5700A (estimated to the nearest 100A only). This is the quarter-wave point: the optical thickness, n_1d_1, is near 5700A/4, or 1425A. The value of the index of refraction of the film is found from Eq. (4):

$$n_1=[n_0(1+(R)^{\frac{1}{2}})/(1-(R)^{\frac{1}{2}})]^{\frac{1}{2}}. \quad (4)$$

In this equation the exact value of n_0, the index of the glass standard, is used. With $R_{\lambda/4}=0.0086$ and $n_0=1.519$, we find $n_1=1.35_2$. Thus an optical thickness $n_1d_1=1425$A (probable error of the order of one or two percent) corresponds to a physical thickness $d_1=1054$A for this film.

Table I gives information for many quarter-wave MgF_2 films similar to that described above for M14. The assumption that the variation of thickness of film with mass placed in the heater is linear is fairly well justified; the mean optical thickness of film per milligram of MgF_2 evaporated is (36.1 ± 0.2)A for the experimental arrangement used.

The calculated indices in column 5 of Table I do not show a consistent trend with wave-length. We might find an "average" for the index of refraction over the region investigated. As most of the films had their minima near 5400A, this might be considered the mean value there. It is 1.35_1. This is significantly lower than the value for the bulk[2] of 1.38; all the samples, in fact, had a calculated index less than 1.38.

Some half-wave MgF_2 films were also prepared. The maxima of reflectivity as a function of wave-length obtained were near the theoretical value, that of the glass alone.

Results for $CaSiO_3$ Films

A quarter-wave film of $CaSiO_3$ on glass gives a maximum of reflectivity as a function of wave-length rather than a minimum as for MgF_2. The absolute value of

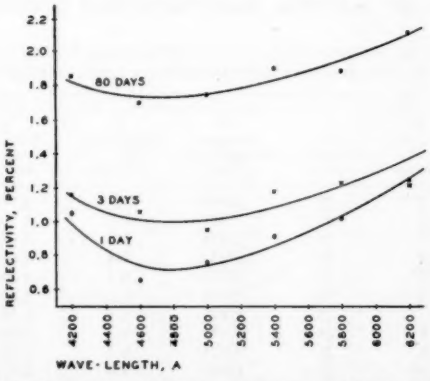

FIG. 4. Aging effects on M1, a quarter-wave MgF_2 film. The curves represent reflectivity measurements carried out 1, 3, and 80 days after preparation of the film.

[2] *Handbook of Chemistry and Physics* (Chemical Rubber Publishing Company, Cleveland, 1947).

$dR/d\lambda$, however, is somewhat less on either side of a $CaSiO_3$ maximum than it is for a MgF_2 minimum. This makes it rather difficult to find the exact position of a maximum. For half-wave films, which give a minimum of reflectivity (theoretically equal to that of the uncoated glass) for $CaSiO_3$, the rate of change of reflectivity is somewhat greater near the extreme value than it is for quarter-wave films. Figure 2 shows sample data for an experimental $CaSiO_3$ film, C9, which had a maximum of reflectivity near 5300A, thus an optical thickness near 1325A, and an index of refraction computed from Eq. (4) of 1.68_5 for 5300A.

Table II gives data for $CaSiO_3$ films analogous to that of Table I for MgF_2. Table II shows a greater

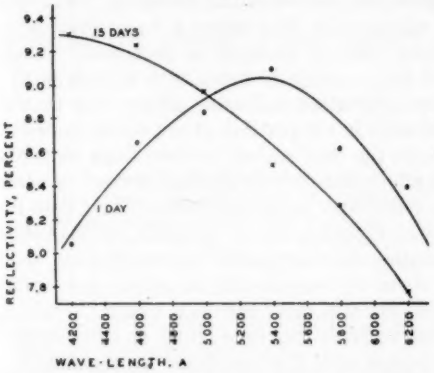

FIG. 5. Aging effects on C9, a quarter-wave $CaSiO_3$ film. The curves represent reflectivity measurements carried out 1 and 15 days after preparation of the film.

fluctuation in optical thickness of film (one-quarter of column 3) per unit of $CaSiO_3$ than does Table I for MgF_2, caused partly, no doubt, by the greater difficulty of evaporation. The mean value is 50.2 ± 0.8 A/mg. In general, the silicate films show a slightly higher instability, a slightly poorer test-immediate-retest correlation than do the fluoride ones. In spite of this, the data of Table II for $CaSiO_3$ hold together better than do those of Table I for MgF_2, and in particular a change of index of refraction with wave-length is evident. The data suggest that the index of refraction changes from about 1.71 near 4400A to about 1.68 near 6000A (see Fig. 3). These are slightly higher than the values of about 1.61–1.65 for sodium light given by Mellor.[3]

Some half-wave $CaSiO_3$ films were also prepared; these will be discussed in the section on aging effects.

Application to the Reduction of Over-All Reflectivity

An interesting practical application of the study of optical properties of thin films is the production of a coated glass of very low reflecting power throughout the visible spectrum. To apply the studies made here, we

[3] J. Mellor, *A Comprehensive Treatise on Inorganic and Theoretical Chemistry* (Longmans, Green and Company, London, 1925), VI, pp. 347ff.

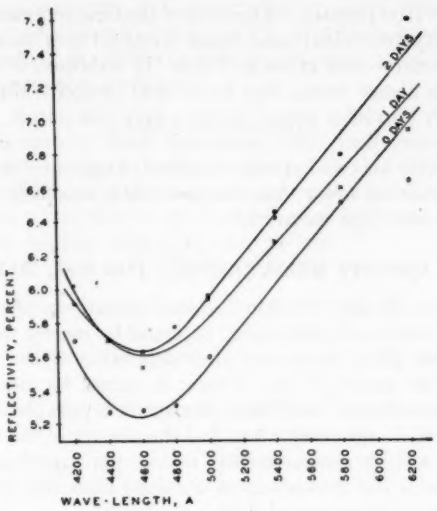

FIG. 6. Aging effects on C11, a half-wave $CaSiO_3$ film. The curves represent reflectivity measurements carried out immediately after preparation of the film, and 1 and 2 days later.

shall use the region 4200A to 6200A. If a single film is used, the position of the quarter-wave minimum should be slightly to the lower wave-length side of the center of the region in order to obtain the smallest over-all reflectivity.

A higher degree of extinction of reflection than that possible with a single film is obtained by combining two films of suitable indices, such as $CaSiO_3$ and MgF_2. One of the possible zeros of reflection occurs if

$$n_2^2 n_0 - n_1^2 = 0, \tag{8}$$

where n_0 is the index of the glass, n_1 that of the first layer (nearest glass), and n_2 that of the second layer (on top), and further the optical thicknesses are both quarter-wave. Condition (8) is nearly met with MgF_2 films of index near $n_2 = 1.35$ and $CaSiO_3$ films of index about 1.69 on glass of index 1.52. Actually (8) holds for only one wave-length; R rises rapidly on either side of the minimum, especially on the side of lower wave-lengths. A smaller over-all reflectivity is obtained if the high index substance (next to glass) has an index lower than that required by Eq. (8).

Table III gives reflectivities for four double films of MgF_2 and $CaSiO_3$, prepared in the laboratory, with optical thicknesses each near quarter-wave. For certain critical thicknesses near quarter-wave a very small change in the thickness of either substance produces a large change in reflectivity. It is apparent that CM4 is drastically lower in reflectivity than any of the rest.

But here an important point has to be considered: the reflectivities in Table III, unlike any of the rest in this chapter, were those present immediately after the glass was removed from the evaporation chamber. (The reflectivities of the other films were measured 24

hours after preparation because of the time necessary for density determinations.) Aging effects set in at once and the reflectivities given in Table III stabilized later at much higher levels. But for a brief period CM4 was nearly "invisible glass;" it was a very dull purple, even in strong light. The theoretical limit of over-all reflectivity for two hypothetical films of optimum indices is somewhat lower than that for CM4, especially near the ends of the spectrum.[4]

IV. DENSITY MEASUREMENTS FOR MgF_2 FILMS

If the density of a thin film were known, the physical thickness could presumably be found by merely weighing the glass before and after deposition of the film. But the density is not known: it cannot be assumed equal to that of the bulk substance. It is possible, however, to do the reverse and find the density of the films from weight measurements, taking the experimental values of the thicknesses as obtained from reflectivity as a function of wave-length.

Such a procedure was carried out successfully for MgF_2. The areas of the glasses used were about 50 cm^2. With thicknesses of the films of the order of 1000A, and densities for MgF_2 films that later came out near 2.5 g/cm^3, the films weighed around two milligrams.[5] Control experiments for moisture equilibrium were run, and numerous weighings of each film made. The weight of a film could be determined finally and reflectivity measured 24 hours after its preparation. Weight measurements and density calculations for six MgF_2 films are shown in Table IV. The mean density is (2.56 ± 0.04) g/cm^3. This is lower than the value of about 3.0 for the bulk substance:[6] evidently the fluoride films are in a very porous state.

V. AGING EFFECTS

It was soon realized through carrying out optical measurements on the same film at different periods after its preparation that its properties may change considerably with time. The films were kept in air, but were protected from dust and mechanical contacts. The rate of change of optical properties with time is usually fairly rapid just after evaporation of a film, but soon drops off, and a quasi-equilibrium state is often reached within two or three days. But weeks and even months later further changes may be noted. Of the total change in three months for a particular film, perhaps one-half may occur within the first three days. The changes consist mainly of two (related) kinds: a change in thickness, and a change in reflectivity.

For silicate films the thickness changes are usually the more pronounced, while for fluoride films the reflectivity changes usually are. Orders of magnitude of the changes that may be expected within the first two or three days after preparation of a film range from one to twenty percent. While the amount of change may vary from film to film, within the experimental errors the direction of change seldom does: the thickness for $CaSiO_3$ becomes less and for MgF_2 either stays the same or decreases slightly; the value of the reflectivity at the quarter-wave extremum goes up for MgF_2 and down for $CaSiO_3$ (except that dispersion effects sometimes enter in the latter case).

Figures 4, 5, and 6 illustrate aging effects for three particular films: M1, a quarter-wave fluoride film that shows a gradual change in reflectivity without much change in the position of the minimum; C9, a quarter-wave silicate film that shows a large drop in optical thickness and an increase in reflectivity (evidently caused by a change of index with wave-length); and C11, an interesting half-wave silicate film that shows little change in the position of the minimum with time (although the time scales for these three films are not comparable), but shows a gradual increase in the value of the reflectivity at the minimum. These data for the particular films are not to be taken as quantitatively representing the changes to be expected for all films, since there is considerable variation, but rather as qualitatively indicating the usual trends. Note that for C11 the value of the reflectivity at the minimum is much higher than that predicted by the theory (4.32 percent for 4600A) and soon rises still higher.

VI. CONCLUSIONS

The values for the indices of refraction of thin films of MgF_2 and of $CaSiO_3$ on glass have been found in this work. The results for the silicate films are sufficiently reliable to establish a change of index of refraction with wave-length: a "normal dispersion" over the region 4200A to 6200A.

$CaSiO_3$ was chosen as one of the two substances to investigate because, of all the glass-like compounds studied before a choice was made (namely, the silicates and sulfates of the alkaline earths), it is the easiest to evaporate reproducibly with the methods used here. Furthermore, the bulk substance is stable and has an index near the optimum for the extinction of reflection with a double film. Actually the film turned out to be less stable than the bulk. Even $CaSiO_3$, too, is rather difficult to evaporate readily, at least as compared with MgF_2.

Perhaps the most important results were obtained in connection with the aging effects observed on thin films. Two definite changes with time can be detected: a lowering of thickness and a change in reflectivity. The thickness diminution is more pronounced for silicate films than for fluoride ones. The change in reflectivity with time is such as to make a maximum or minimum of reflectivity as a function of wave-length

[4] See, for example, R. B. Muchmore, J. Opt. Soc. Am. **38**, 20 (1948).

[5] A. E. Conrady, Proc. Roy. Soc. London **101**, 211 (1922), described procedures for weighing masses as large as 50 grams with probable errors of 0.001 mg or less.

[6] See reference 3, IV, pp. 296f.

less extreme; near a maximum the reflectivity drops with time and near a minimum it rises. Such observations suggest that for some time after deposition of a film the molecules may diffuse and perhaps crystallize, so that while the average thickness may remain the same, the surface of the film will be uneven.[7] Since the light beam used for reflectivity measurements covers an area of the films very large as compared with molecular dimensions, the uneven surface would result in just those changes of reflectivity with time actually observed. It would be a leveling-off process of reflectivity as a function of wave-length.

[7] This idea (applied to a different situation) is discussed by R. G. Picard and O. S. Duffendack, J. App. Phys. 14, 291 (1943).

ACKNOWLEDGMENTS

I wish to thank Professor Hiram W. Edwards for considerable assistance and advice, especially with experimental techniques. Professor Harvey E. White, too, offered valuable suggestion and opinions.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 39, NUMBER 10 OCTOBER, 1949

Remarks on the Classical and Quantum-Mechanical Treatment of Partial Polarization*

U. FANO
National Bureau of Standards, Washington, D. C.
(Received June 10, 1949)

The methods for the convenient treatment of partial polarization are discussed. In particular, it is shown (1) how to introduce the "Stokes parameters" and the "Poincaré sphere" representation in a direct phenomenological way; (2) how an idealized experimental procedure may illustrate the physical analysis of partial polarization; (3) how the Stokes parameter representation fits naturally into the quantum theory of radiation; (4) how the Stokes method applies to Compton scattering.

THE proper treatment of problems involving partially polarized electromagnetic radiation frequently does not appear as obvious as it should be. A suitable analytical tool for this treatment was indeed provided by Stokes as early as 1852.[1] Stokes himself showed a great insight in the physical significance of his method, but classical physics was, usually, somewhat obscure about the superposition of incoherent waves which plays a central role in the relationship between partially and wholly polarized light. Quantum mechanics has now clarified the meaning of incoherent superposition. Still, no clear treatment of partial polarization has yet found its way into the standard textbooks or has become a part of general physical knowledge.

The Stokes method has, however, found increasing application in recent years,[2] and its significance and application have been the object of much attention by Mueller and his co-workers.[3] The purpose of the present paper is to complement this recent work by presenting some considerations that should help in visualizing the proper treatment of partial polarization effects. These considerations concern: (1) The direct phenomenological introduction of the Stokes parameters in connection with the "Poincaré sphere" representation, (2) an idealized experimental procedure that illustrates the physical content of alternate descriptions of partially polarized radiation, (3) the treatment of polarization in the quantum theory of radiation, (4) the application of the Stokes method to Compton scattering.

The language and the tools of optics will be used in the following discussion for convenience, but the arguments presented can be applied, in principle, to any electromagnetic ratiation.

1. INTRODUCTION OF THE STOKES PARAMETERS

Consider first the elementary case of linearly polarized light, whose analysis can be performed by a simple Nicol prism. If a beam of light is completely linearly polarized, the fraction of it which is transmitted by a Nicol prism is

$$\cos^2\theta = (1+\cos 2\theta)/2,$$

where θ is the angle between the direction of polarization of the incident light and the direction of polarization of the light that would be fully transmitted by the prism. If the light is only partially linearly polarized, the fraction transmitted can be represented as:

$$(1+P\cos 2\theta)/2. \qquad (1)$$

Here P is the "degree of polarization," which is determined by observing the maximum fraction trans-

* Work supported in part by an ONR contract.

[1] G. G. Stokes, *Mathematical and Physical Papers* (Cambridge, University Press, London, 1901), Vol. 3, p. 233, or Trans. Camb. Phil. Soc. 9, 399 (1852).

[2] See, e.g., P. Soleillet, Ann. de physique 12, 23 (1929); F. Perrin, J. Chem. Phys. 10, 415 (1942); S. Chandrasekhar, Astrophys. J. 104, 110 (1946); 105, 424 (1947).

[3] H. Mueller, *Foundations of Optics* (in press). N. G. Parke III, M.I.T., Res. Lab. Electr., Tech. Rep. 70 (June 30, 1948). R. Clark Jones, J. Opt. Soc. Am. 37, 107 (1947).

mitted (i.e., $(1+P)/2$ for $\theta=0$) or the minimum fraction (i.e., $(1-P)/2$ for $\theta=90°$).

In general, light may be expected to be elliptically polarized. The quantitative analysis of polarization can then be completely performed by means of the single arrangement of a $\lambda/4$ plate and Nicol prism shown schematically in Fig. 1. (This point was also emphasized by Stokes[4] but is not generally mentioned in textbooks.) The fraction of an incident beam of light which is transmitted by this analyzer depends on its setting, that is, on the orientation of the $\lambda/4$ plate and of the

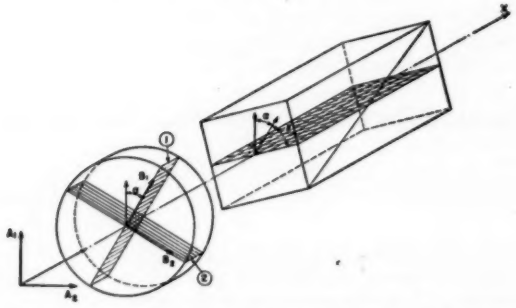

FIG. 1. Diagram of an ideal complete polarization analyzer consisting of a quarter-wave-length plate and a Nicol prism. Light propagates "slow" through the $\lambda/4$ plate when the electric field is in plane (1), "fast" when in plane (2); it is admitted by Nicol when in cross-hatched plane. The setting of the analyzer is characterized by the pair of angles α and β or by a single complex vector $\mathbf{B}=\mathbf{B}_1\cos\beta+i\mathbf{B}_2\sin\beta$.

Nicol prism. Unless the polarization is circular, there is a single setting for which the fraction transmitted is a maximum, and another setting for which it is a minimum, as in the case of linear polarization. The mean of these two fractions is just $\frac{1}{2}$ and therefore these fractions can be indicated by $(1+P)/2$ and $(1-P)/2$ as in the case of linear polarization.

In the setting of maximum transmission, the $\lambda/4$ plate has the effect of transforming an initial elliptical polarization into a linear one. If then the $\lambda/4$ plate is kept fixed and the Nicol is rotated, the transmission will vary according to (1), that is, it will decrease to $(1-P)/2$ and then return to $(1+P)/2$; the angle θ stands here for the difference $\beta-\beta_0$, where β_0 is the value of β at optimum setting. If, on the contrary, the $\lambda/4$ plate and the Nicol are rotated together, that is, if α is varied, the transmission still decreases and then increases sinusoidally but the peak variation depends on the kind of polarization; it is a maximum for linear polarization (i.e., for $\beta_0=0$) and it vanishes for circular polarization (i.e., for $\beta_0=\pm 45°$).

In general, the transmission obtained for any setting of the analyzer is represented by (1) with

$$\cos 2\theta=\cos 2\beta\cos 2\beta_0\cos 2(\alpha-\alpha_0)+\sin 2\beta\sin 2\beta_0.$$

This is a familiar formula from spherical trigonometry which determines the arc of great circle 2θ between two points on a sphere whose latitudes and longitudes are $(2\beta, 2\alpha)$ and $(2\beta_0, 2\alpha_0)$. It is therefore convenient to follow Poincaré's method of mapping all settings of the analyzer on a sphere according to this rule. Each polarization can then be similarly mapped at the point $(2\beta_0, 2\alpha_0)$ corresponding to the setting of maximum transmission. The $\cos 2\theta$ can then be represented as the scalar product of two vectors of unit length pointing from the center of the sphere to the points $(2\beta, 2\alpha)$ and $(2\beta_0, 2\alpha_0)$.

A further simplification is achieved by choosing this second vector, not of unit length but of length P, so that it represents not only the quality but also the degree of the polarization. If this vector is indicated by $\mathbf{P}$, while the unit vector which points to $(2\beta, 2\alpha)$ and represents the analyzer setting is indicated by $\mathbf{Q}$, the transmission formula (1) can be written as

$$(1+\mathbf{P}\cdot\mathbf{Q})/2. \tag{1'}$$

Stokes pointed out that the retarding plate in the analyzer of Fig. 1 does not have to be exactly $\lambda/4$. If it is not, no setting of the analyzer will yield complete transmission or extinction of circularly polarized light, that is, $\mathbf{Q}$ cannot be brought at or near the poles of the Poincaré sphere. It should further be noted that any device whose response is proportional to the intensity of an incident beam of light and depends on the polarization of the light can be used as a polarization analyzer. The length of the vector $\mathbf{Q}$ which characterizes the setting of such an analyzer is, however, in general, ≤ 1.

The theory of the analysis of polarization leading to (1') is straightforward. The electric field of an incident polarized monochromatic beam may be represented by

$$\mathbf{E}=(\mu_0/4\epsilon_0)^{\frac{1}{4}}\{\mathbf{A}\exp[2\pi i(kx-\nu t)]+\text{compl. conj.}\}, \tag{2}$$

where $\mathbf{A}$ is a complex vector that describes the intensity and the polarization of the light and the other symbols have their usual meaning. The intensity of the incident light is $\mathbf{I}=|\mathbf{A}|^2$. The intensity of the light transmitted by the analyzer is simply $|\mathbf{A}\cdot\mathbf{B}|^2$, where the vector $\mathbf{B}$ is defined in Fig. 1. If $\mathbf{A}$ is expressed by means of the vectors $\mathbf{A}_1$ and $\mathbf{A}_2$ of Fig. 1:

$$\mathbf{A}=a_1\mathbf{A}_1+a_2\mathbf{A}_2, \tag{2'}$$

we have

$$\begin{aligned}|\mathbf{A}\cdot\mathbf{B}|^2=|a_1(\cos\alpha\cos\beta-i\sin\alpha\sin\beta)+a_2(\sin\alpha\cos\beta&\\+i\cos\alpha\sin\beta)|^2=(\tfrac{1}{2})[(|a_1|^2+|a_2|^2)&\\+(|a_1|^2-|a_2|^2)\cos 2\alpha\cos 2\beta&\\+(a_1^*a_2+a_2^*a_1)\sin 2\alpha\cos 2\beta&\\+i(a_1^*a_2-a_2^*a_1)\sin 2\beta].&\end{aligned} \tag{3}$$

The first term in the brackets is the intensity I, the remaining terms constitute the expression of the scalar product of two vectors, namely $I\mathbf{P}$ and $\mathbf{Q}$. This is to say that, in agreement with (1'), Eq. (3) can be written

[4] See reference 1, *Mathematical and Physical Papers*, p. 197.

in the form:

$$|\mathbf{A}\cdot\mathbf{B}|^2=I(1+\mathbf{P}\cdot\mathbf{Q})/2.^5 \qquad (3')$$

The electric field (2) corresponds to a completely polarized wave; in agreement with this it can be verified from (3) that $|I\mathbf{P}|^2=I^2$, i.e., $P=1$. If, however, two completely polarized waves with different vectors $\mathbf{A}'$ and $\mathbf{A}''$ were superposed incoherently and filtered through the analyzer, the transmitted intensity would be

$$|\mathbf{A}'\cdot\mathbf{B}|^2+|\mathbf{A}''\cdot\mathbf{B}|^2=I'(1+\mathbf{P}'\cdot\mathbf{Q})/2 +I''(1+\mathbf{P}''\cdot\mathbf{Q})/2=I(1+\mathbf{P}\cdot\mathbf{Q})/2,$$

with

$$I=I'+I'', \quad I\mathbf{P}=I'\mathbf{P}'+I''\mathbf{P}'', \quad P\leq 1. \qquad (4)$$

The intensity I and the three components of the vector $I\mathbf{P}$ (expressed in any system of coordinates attached to the Poincaré sphere) represent the four "Stokes parameters" describing the intensity and polarization of a beam of light. Equation (4) represents the addition theorem of Stokes parameters when incoherent light beams are superposed.

2. ALTERNATE DESCRIPTIONS OF PARTIALLY POLARIZED LIGHT

Equation (4) applies also to the analysis of partially polarized light into two (or more) incoherent completely polarized components. The quantities I and $\mathbf{P}$ are then regarded as known, and I', $\mathbf{P}'$, I'', $\mathbf{P}''$ are chosen so as to fulfill (4) and the complete polarization conditions $P'=P''=1$. The construction in Fig. 2 shows that each choice of the vectors $I'\mathbf{P}'$, $I''\mathbf{P}''$ corresponds to the choice of one point of an ellipsoid. A well-known special choice is that of $\mathbf{P}'=-\mathbf{P}''=\mathbf{P}/P$, in which the component beams are polarized perpendicularly to each other. (The Poincaré mapping involves *twice* the geometrical angle between the planes of polarization.)

The physical significance of alternate decompositions of partially polarized light into completely polarized components can be illustrated by the idealized experimental procedure described in Fig. 3. It deals with the polarization analysis of the light which is emitted by atoms whose mechanical angular momentum has been preliminarily fixed by filtration through a Stern-Gerlach molecular beam device. If the light is filtered through the analyzer of Fig. 1 and then detected by a photocounter and scored by the "total recorder," the light is not found to be completely polarized (unless the total quantum number j' of the atomic state after light emission is zero). The procedure may then be varied by filtering the atoms after light emission through a second Stern-Gerlach device which is set to accept only atoms of magnetic quantum number m'. The filtered atoms are then detected by a counter and the coincidences between this counter and the photo-counter may be recorded. The observation is thereby confined to light emitted by atoms in a transition between states of *known quantum numbers* (j, m) and (j', m'). This light is then found to be *completely polarized*. The partially polarized light scored by the total recorder can thus be regarded as the superposition of incoherent polarized components emitted by atoms that land in states with different values of the quantum number m'. However, this analysis is still arbitrary since it can be performed in different ways by changing the direction of the magnetic field in the second Stern-Gerlach device, that is, by selecting the quantum numbers m' with respect to different axes of quantization.

In conclusion, *complete* polarization of light is observed when the orientation of the radiating atoms is subjected to a *maximal* set of observations before and after emission. Conversely, *incomplete* polarization is associated with *incomplete* determination of the mechanical state of the light source. This result, which is rather obvious from the standpoint of quantum mechanics, can be formulated more generally by stating that incompletely polarized light is light that has been emitted by—or has interacted with—an incompletely analyzed mechanical system.

3. QUANTUM-MECHANICAL DISCUSSION

The quantum theoretical treatment of electromagnetic radiation fits in very well with the treatment of optics by means of the Stokes parameters. This is quite natural since the Stokes parameters are actually the "observable" quantities in phenomenological optics.

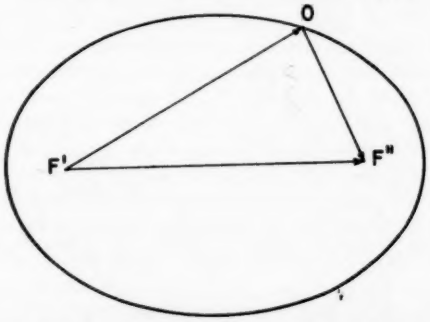

FIG. 2. Diagram of the composition of polarizations $F'O=I'\mathbf{P}'$, $OF''=I''\mathbf{P}''$, $F'F''=I\mathbf{P}$. The point O must be on an ellipsoid since the sum of its distances from F' and F'' (the foci) is subject to the condition $F'O+OF''=I'+I''=I$.

According to quantum theory, the "state" of electromagnetic radiation can be described by a wave function whose variables are the numbers of photons in each of the quantum states of a complete set. A complete set of quantum states includes for each frequency and direction of propagation, two states of opposite polarization. Light in one of these states is freely admitted by a polarization analyzer characterized by a vector $\mathbf{Q}$ (see Section 1), while radiation of the other state is wholly rejected by the same analyzer. Alternate com-

[5] The complex plane vectors $\mathbf{A}$ and $\mathbf{B}$ can be regarded as spinors; Eq. (3') is then a familiar formula from spinor algebra.

plete sets of quantum states correspond to different choices of $\mathbf{Q}$.

Pauli[6] has treated the transformations which the wave function undergoes when the choice of $\mathbf{Q}$ is changed. When there are N photons of given frequency and direction of propagation, the wave function has only two variables N_1 and N_2, namely, the numbers of photons in the two opposite polarization states, with $N_1+N_2=N$. When $\mathbf{Q}$ is changed, that is, when its representative point on the Poincaré sphere is moved over the sphere, this wave function transforms, as the wave function of an atomic system with angular momentum $(N/2)\hbar$ does when the space coordinates are rotated. If the light is completely polarized, there is one setting of the analyzer which admits all the light freely. The corresponding wave function is

$$\varphi(N_1, N_2)=\delta_{NN_1}\delta_{0N_2} \tag{5}$$

(where δ_{ik} equals 0 or 1 depending on whether its indices are different or equal). This setting of the analyzer is that for which the reference vector $\mathbf{Q}$ coincides with the polarization vector $\mathbf{P}$ of the light. The wave function (5) of a group of N completely polarized photons transforms with respect to changes of $\mathbf{Q}$ as the wave function $\psi_{N/2,N/2}$ of an atomic system does with respect to rotations of space. (The first index of ψ represents the total angular momentum quantum number and the second index represents the magnetic quantum number with respect to $\mathbf{P}$ as axis of quantization.)

According to (1′), an analyzer whose representative vector is $\mathbf{Q}$ admits a fraction $(1+\mathbf{P}\cdot\mathbf{Q})/2$ of the total intensity of light whose polarization vector is $\mathbf{P}$. In the quantum-mechanical formulation one considers the "*mean* fraction of photons" admitted instead of the "fraction of the intensity." The quantum-mechanical operator corresponding to the classical polarization vector $\mathbf{P}$ is expressed by means of the operator $\mathbf{J}$ which transforms the wave function φ as an infinitesimal rotation of the polarization mapping on the Poincaré sphere. It follows from the transformation rule for φ indicated above that the mean value of $\mathbf{J}$ is just $N\mathbf{P}/2$ (in fact,

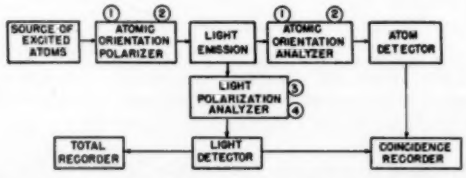

FIG. 3. Diagram of system for ideal simultaneous analysis of polarization of atoms and of the light they emit. An "atomic orientation polarizer" or "analyzer" is a Stern-Gerlach device which accepts only atoms with a selected value of the magnetic quantum number referred to a selected axis of quantization. (1) Setting of axis of quantization. (2) Setting of quantum number selector. (3) Setting of $\lambda/4$ plate as in Fig. 1. (4) Setting of Nicol as in Fig. 1.

[6] W. Pauli, *Handbuch der Physik* (Verlag. Julius Springer, Berlin 1933), Vol. 24/1, p. 252. See also G. Araki, Phys. Rev. 74, 672 (1948).

an infinitesimal rotation by an angle ϵ around the direction of $\mathbf{P}$ has the effect of multiplying the function (5) by $1+\epsilon N/2$, since it transforms as $\psi_{N/2,N/2}$. Therefore, the mean fraction of photons admitted by an analyzer represented by $\mathbf{Q}$ is the mean value of the operator:

$$[1+(2/N)\mathbf{J}\cdot\mathbf{Q}]/2. \tag{6}$$

The mean *number* of photons admitted is more simply the mean value of:

$$N/2+\mathbf{J}\cdot\mathbf{Q}. \tag{7}$$

Here N itself can be considered as an operator, namely, that which is indicated, in Pauli's notation,[5] as

$$\sum_{\lambda=1,2} a^*(\lambda)a(\lambda).$$

The same result (7) can be obtained by calculating the intensity transmitted by the analyzer according to (3), i.e., according to classical optics, and *then* taking into account that the amplitudes a_1 and a_2 must be treated as operators. These amplitudes differ from Pauli's operators only by a normalization factor, and — to within this same factor — $I=|a_1|^2+|a_2|^2$ coincides with the operator N while the components of $I\mathbf{P}$ coincide with the operator $2\mathbf{J}$. (Care must be taken in fixing the order of starred and unstarred operators.)

At this point it is no longer necessary to consider the special case where there is a fixed number of photons, once it is understood that N is to be treated as an operator in determining the mean value of (7). The extension of (7) to cover the cases of non-monochromatic light and of light flowing in various directions should be possible, by carrying over the classical method studied respectively by Parke[3] and Soleillet.[2]

Partially polarized light will generally be described by the incoherent superposition of different wave functions, in keeping with the fact that partial polarization means incomplete knowledge of the "state" of the radiation. There exist, however, completely defined states which cannot be brought into the form (5) and therefore represent a state of partial polarization. An example of such states is seen in the wave function

$$\varphi(N_1, N_2)=\delta_{N_1,N'}\delta_{N_2,N-N'} \tag{8}$$

with $N'\neq N$, which represents a state with N' photons polarized one way and $N-N'$ polarized the opposite way. With reference to the discussion of Section 2, it appears that a state of this type may arise when it is known that N' photons were emitted by atoms with a certain final orientation, and the remaining photons by atoms of different final orientation, but no score is kept of the correlation between individual photons and atoms.

4. APPLICATION TO COMPTON SCATTERING

The Klein-Nishina theory serves to predict at least in principle not only the intensity but also the polarization of Compton scattered x-rays. Nishina himself first

described the radiation scattered by an electron as consisting of two incoherent elliptically polarized waves.[7] This method of representing a partial polarization is not very convenient, especially for application to the study of multiple scattering.

The cross section for Compton scattering calculated by more recent methods[8] is seen to depend on "the angle between the directions of polarization of the incident and scattered radiation." This statement means that the cross section determines the probability for a plane-polarized photon to be scattered in a certain direction *and then to pass* through a hypothetical filter which accepts only radiation polarized in a certain plane. Elliptical polarization is not considered in that treatment.

The Stokes method proves convenient for dealing with partially polarized x-rays, especially in the course of multiple scattering and it has been applied to this problem by Wightman.[9] According to this method, the Klein-Nishina cross section is expressed in the form of a 4×4 matrix which relates the four Stokes parameters of the incident and scattered radiation. This expression will be presented here in a form equivalent to, but slightly more general than, the one presented by Wightman.

If N photons are scattered in a certain direction and their polarization is represented by $\mathbf{P}$, the number accepted by a polarization filter with a setting $\mathbf{Q}$ is $N_Q=N(1+\mathbf{P}\cdot\mathbf{Q})/2$. The Stokes parameters N and $N\mathbf{P}$ of the scattered radiation are related linearly to the parameters N_0 and $N_0\mathbf{P}_0$ of the incident radiation by a matrix $\mathbf{T}=(t_{ik})$; this relationship may be written in the form:

$$(N, N\mathbf{P})=\mathbf{T}\cdot(N_0, N_0\mathbf{P}_0), \quad \text{or} \quad u_i=\sum_{k=0}^{3} t_{ik}v_k, \qquad (9)$$

where (u_0, u_1, u_2, u_3) stands for $(N, N\mathbf{P})$ and (v_0, v_1, v_2, v_3) stands for $(N_0, N_0\mathbf{P}_0)$. The combination $(1, \mathbf{Q})$ may be similarly indicated by (w_0, w_1, w_2, w_3), so that:

$$N_Q=(\tfrac{1}{2})\sum_{i,k=0}^{3} t_{ik}w_iv_k=(1, \mathbf{Q})\cdot\mathbf{T}\cdot(N_0, N_0\mathbf{P}_0). \qquad (10)$$

The Klein-Nishina differential cross section $d\sigma/d\Omega$ gives the ratio of N_Q to $N_0=v_0$, when N_Q represents photons per unit solid angle and N_0 represents photons per square centimeter. The cross section depends on $\mathbf{P}_0$ and $\mathbf{Q}$. This dependence is brought out explicitly by expressing the relationship between N_Q and N_0 in the form (10). $\mathbf{T}$ represents then the matrix form of the Klein-Nishina cross section.

The explicit form of $\mathbf{T}$ has been obtained by the method used by Wightman, with the minor change of assuming a specific instead of an average direction $\mathbf{S}$ for the initial spin of the scattering electron. The directions of the incident and scattered photon are indicated by $\mathbf{n}_0$ and $\mathbf{n}$, their momenta by $\mathbf{k}_0$ and $\mathbf{k}$ in units of mc, and the angle of scattering between $\mathbf{k}_0$ and $\mathbf{k}$ by ϑ. It is necessary to use some specific system of coordinates attached to the Poincaré spheres for the incident and scattered radiation; in our formula the Cartesian axis 1 corresponds to linear polarization in the plane of scattering, axis 2 to linear polarization in a plane 45° to the right, and axis 3 to right-handed circular polarization. With these notations, the expression of $\mathbf{T}$ is:

$$T=(\tfrac{1}{2})(e^2/mc^2)^2(k/k_0)^2 \times \begin{bmatrix} 1+\cos^2\theta+(k_0-k)(1-\cos\theta) & -\sin^2\theta & 0 & (1-\cos\theta)(\mathbf{k}_0\cos\theta+\mathbf{k})\cdot\mathbf{S} \\ -\sin^2\theta & 1+\cos^2\theta & 0 & (1-\cos\theta)\mathbf{n}\times\mathbf{n}_0\cdot\mathbf{k}_0\times\mathbf{S} \\ 0 & 0 & 2\cos\theta & (1-\cos\theta)\mathbf{k}\times\mathbf{n}_0\cdot\mathbf{S} \\ (1-\cos\theta)(\mathbf{k}\cos\theta+\mathbf{k}_0)\cdot\mathbf{S} & (1-\cos\theta)\mathbf{n}_0\times\mathbf{n}\cdot\mathbf{k}\times\mathbf{S} & (1-\cos\theta)\mathbf{k}_0\times\mathbf{n}\cdot\mathbf{S} & 2\cos\theta+(k_0-k)(1-\cos\theta)\cos\theta \end{bmatrix}.$$

The upper left matrix element is the one which relates the intensity of the scattered to that of the incident radiation and is just the usual average differential cross section for scattering by an angle ϑ irrespective of the polarization; [notice that $k_0-k=k_0^2(1-\cos\vartheta)/1+k_0(1-\cos\vartheta)$]. The second term of the first row gives rise to a departure of the scattered intensity from cylindrical symmetry. This effect is proportional to the degree of linear polarization of the incident radiation; its importance decreases for hard radiation since the matrix element involved does not increase with k_0. The whole 3×3 submatrix on the upper left side can be derived directly from the standard formula for the cross section, while the remaining elements represent the effects of elliptical or circular polarization. The non-diagonal elements on the lower and right edge of the matrix average out when the radiation scattered by electrons of different spin orientation is superposed incoherently. No elliptical polarization arises in this case unless the incident radiation itself has a degree of elliptical polarization.

I wish to thank Mr. P. R. Karr for discussing the problem and checking the calculations.

[7] Y. Nishina, Zeits. f. Physik **52**, 869 (1928).
[8] See e.g., W. Heitler, *Quantum Theory of Radiation* (Oxford University Press, London, 1944), p. 154, Eq. (47).
[9] A. Wightman, Phys. Rev. **74**, 1813 (1948).

A Method of Measuring the Contrast Rendition of Telescopic Systems*

HOWARD S. COLEMAN, GEORGE W. ARNOLD, JR., AND WALTER D. LUEDECKE
The Optical Research Laboratory, The University of Texas, Austin, Texas

(Received June 29, 1949)

A method has been developed by means of which the contrast rendition of telescopic systems can be measured using collimated light. The method utilizes an optical-electronic apparatus referred to as "the contrast rendition apparatus model 2" and is designed following the basic principles for such equipment as described in a previous paper. This apparatus is 60 inches long, 30 inches wide, and 65 inches high, and includes an assembly referred to as "the artificial sun." This assembly permits the measurement of the contrast rendition of telescopic systems for the simulated condition of observation in which a bright luminous body may appear to be located near the region under examination. The contrast rendition has been measured for a number of prism and lens erecting telescopic systems for a variety of conditions simulating situations of intended uses for such optical systems. The numerical values of the contrast rendition were found to range from 0 to 93 percent for the optical systems studied.

INTRODUCTION

A METHOD has been developed for measuring the contrast rendition of telescopic systems in collimated light. The contrast rendition is a property inherently dependent upon the particular design of a given optical system and its manner of use. This quantity is a measure of the fidelity of an image of an object formed by an optical system with respect to brightness contrast and is perhaps one of the most important properties of an optical device. For common optical systems, contrast rendition values have been found to range from 0 to 100 percent depending upon the design and conditions under which they are used. Contrast rendition values of less than 100 percent indicate the existence of a certain amount of non-image-forming light which has become mixed in an optical system with image-forming light. This non-image-forming light reduces the brightness contrast of images formed by optical systems and hence impairs the "visibility" of objects viewed with such systems. In addition it appears to augment falsely the transmittance of optical systems. Practically all optical systems have contrast rendition values less than 100 percent.

It is the purpose of this paper to describe a practicable method by means of which the contrast rendition can be measured for telescopic systems under simulated conditions of their intended uses.

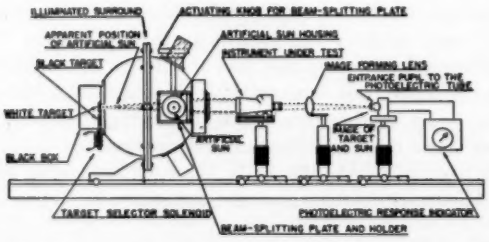

FIG. 1. Schematic diagram of contrast rendition apparatus model 2.

THE APPARATUS

The apparatus described in this paper to measure the contrast rendition of telescopic systems has been designated as "the contrast rendition apparatus model 2." This designation has been made following that of an apparatus previously described,[1] which in general, required more space than would ordinarily be available in most research and production centers for housing such equipment. The earlier equipment required a darkened room having a length of 50 feet, a width of 25 feet, and a height of 15 feet. The contrast rendition apparatus model 2 is of "table" size having a length of 60 inches, a width of 30 inches, and a height of 65 inches, and does not require a darkened room.

The contrast rendition apparatus model 2 is an optical-electronic system by which the brightness contrast of the image of an object formed by an optical system under test can be measured and compared with the inherent contrast of the object. When this comparison is made on a percentage basis, a quantity, referred to as the contrast rendition, is obtained. For most practical purposes a "black" object viewed against a "white" background is used, thereby resulting in an inherent contrast of unity for the object. This choice of inherent contrast for the object requires only that the contrast of its image be measured in order to compute the contrast rendition of an optical system under test. As shown in Figs. 1 and 2, the test object used in the contrast rendition apparatus model 2 is usually a black circular hole in a hollow spherical shell which is painted white inside and illuminated. This test object is referred to as the "target." The image of the target formed by an optical system under test is projected on the entrance pupil leading to a photoelectric tube as shown in Fig. 1. The apparatus is adjusted such that if the image has the same brightness contrast as the object, effectively no photoelectric response is obtained (for a black target). If there is a certain amount of non-image-forming light mixed with the image-forming light, an

* The method described in this paper was developed in connection with Task 1 of Contract NOrd-10403 between the Bureau of Ordnance of the Navy Department and The University of Texas.

[1] Howard S. Coleman, *Stray light in optical systems*, J. Opt. Soc. Am. 37, 434–451 (1947).

effective photoelectric response is indicated, the magnitude of which is a measure of the degree of imperfection of the contrast rendition (i.e., less than 100 percent) of the optical system under consideration.

For purposes of description, the contrast rendition apparatus model 2 may be regarded as consisting of

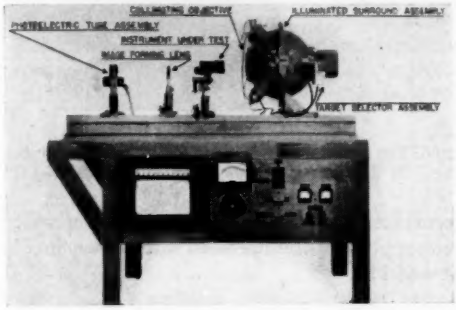

FIG. 2. The contrast rendition apparatus model 2.

the following components:

1. The target
2. The surround
3. The collimating objective
4. The artificial sun assembly
5. The electronic assembly
6. The image-forming lens
7. The photoelectric assembly
8. The photoelectric response indicators

The Target

The target used in the contrast rendition apparatus model 2 is usually a black hole at the vertex of a white illuminated hemispherical surround. The subtense of the target may be varied from one tenth of a degree to ten degrees. Provision has been made for covering the "black hole" with various shades of processed photographic paper thereby changing the inherent contrast of the target with respect to the background against which it is viewed. The size of the target used in the contrast rendition apparatus depends upon the magnification of the optical system under test. The higher the magnification, the smaller is the subtense of the target required. For many optical systems of interest, the target subtense required is approximately one quarter of a degree.

The electrical and mechanical mechanisms associated with the target are shown in Fig. 3. A "paddle"-shaped assembly actuated by electromagnetic relays is used to remove the target from the field of view and to complete the white surround by "filling in" the hole in the surround by a white inset. This process makes it possible to refer all the brightness measurements required in using the contrast rendition apparatus to the brightness level of the surround without rotating the optical system of the contrast rendition apparatus off-axis. In addition, provision has been made to include an irregularly shaped black target for focusing purposes. This special target is sometimes required when the general illumination on the entrance pupil of the photoelectric tube used in the contrast rendition apparatus is small (from a visual point of view).

The Surround

The background against which the target is viewed in the contrast rendition apparatus model 2 is the inside of an illuminated white spherical surface arranged as shown in Figs. 1 and 4. The spherical surface is referred to as "the surround."

The surround is illuminated by four incandescent tungsten lamps operated from a regulated power supply. The voltage at which the lamps are operated is controlled by means of a variac transformer. The surround is used to simulate various brightness conditions of the background against which objects may be viewed using optical devices.

The Collimating Objective

The contrast rendition apparatus model 2 is provided with an optical system whereby the target is viewed in collimated light. This arrangement utilizes a telescopic objective, referred to as "the collimating objective," having a clear aperture of 3 inches and an effective focal length of 15.75 inches. The collimating objective is adjusted such that the target lies close to its principal focus. The use of this arrangement makes it possible to simulate field conditions with respect to the focus of optical systems under test.

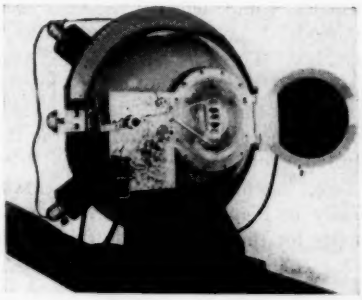

FIG. 3. The target selector assembly of the contrast rendition apparatus model 2.

The Artificial Sun Assembly

In order to provide a means of simulating the conditions of observation in which the target might be located somewhere within the vicinity of the sun, an assembly referred to as "the artificial sun" has been incorporated into the contrast rendition apparatus model 2. The artificial sun assembly consists of a 25-watt Western Union

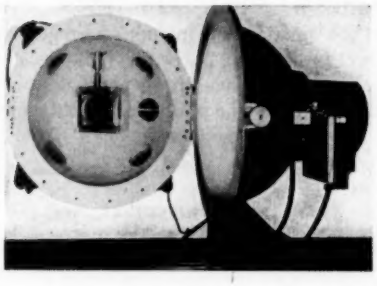

FIG. 4. The illuminated surround and beam-splitting plate of the contrast rendition apparatus model 2.

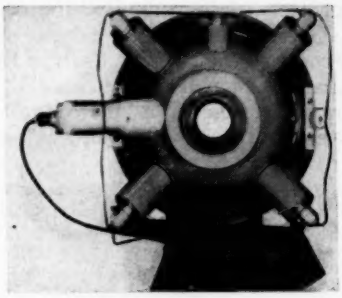

FIG. 6. The artificial sun assembly and collimating objective of the contrast rendition apparatus model 2.

arc lamp subtending approximately one half of a degree with respect to the collimating objective mentioned above. The assembly includes the use of a beam-splitting plate by means of which the image of the arc lamp mentioned can be made to appear at any desired angular separation from the target. The beam-splitting plate is a pellicle having a transmission factor of approximately 92 percent and a reflection factor of approximately 8 percent when used at an angle of incidence of 45°. The beam-splitting plate and the artificial sun assembly are shown in Figs. 4, 5, and 6. The beam-splitting plate is oriented at approximately 45° with respect to the mechanical axis of the arc lamp and approximately at 45° to the optical axis of the collimating system. Provision has been made to rotate the beam-splitting plate through known angles thereby causing the image of the arc lamp to appear at different distances from the target as desired.

The Electronic Assembly

The electrical components of the contrast rendition apparatus model 2 have been designated as "the electronic assembly." These may be regarded as consisting of three parts. The first is the voltage regulation unit and its control for the surround, the second is the power supply for the photoelectric tube, and the third is the power supply for actuating the relays used in connection with the target assembly. These components are shown schematically in Fig. 7. The electrical circuits of the power supply for the photoelectric tube together with performance curves, showing the range over which the power supply may be operated, are shown in Figs. 8, 9, 10, and 11.

The Image Forming Lens

In order to be able to make measurements of the magnitude of the non-image-forming light in optical systems under conditions in which they are generally used, provision has been made in the contrast rendition apparatus model 2 for testing such optical systems that may be focused from infinity to relatively short distances. This provision includes the use of an optical component referred to as "the image forming lens." This lens is an achromatic doublet and is located at the exit pupil of the optical system under test. In many cases, experience has shown that the same results are obtained with or without the use of this component.

The Photoelectric Assembly

The arrangement of the electrical and mechanical parts associated with the photoelectric tube have been designated as "the photoelectric assembly." This assembly consists primarily of a photoelectric tube and its associated electrical and mechanical parts, as shown in Figs. 1 and 2. The photoelectric tube is enclosed in a light-tight shield and is arranged such that only light passing through an aperature, referred to as "the entrance pupil to the photoelectric tube," strikes the

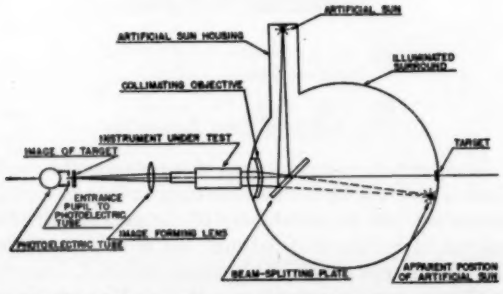

FIG. 5. Optical diagram for contrast rendition apparatus model 2.

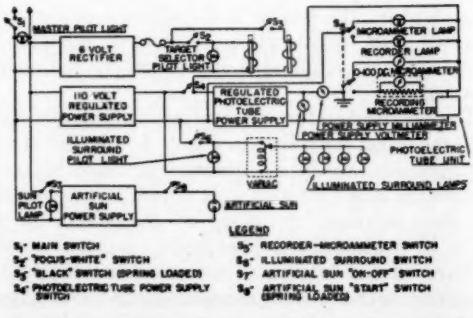

FIG. 7. Wiring diagram for contrast rendition apparatus model 2.

photo-sensitive element of the tube. The entrance pupil to the photoelectric tube may be varied from approximately $\frac{1}{4}$ of a millimeter to 20 millimeters as may be desired. The electrical components of the photoelectric assembly are shown in Figs. 7, 8, and 9.

The Photoelectric Response Indicators

Two units have been provided for indicating the photoelectric responses obtained in the contrast rendition apparatus model 2. The first of these is a recording type microammeter and the second is a conventional

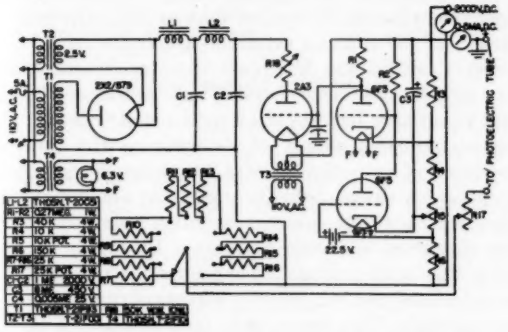

FIG. 8. Wiring diagram for the photoelectric tube power supply

non-recording unit. The recording type of indicator has been found to be desirable in making contrast rendition measurements in subdued lighting and provides a valuable check on the data recorded by the operator. The principal use for the non-recording unit is in the

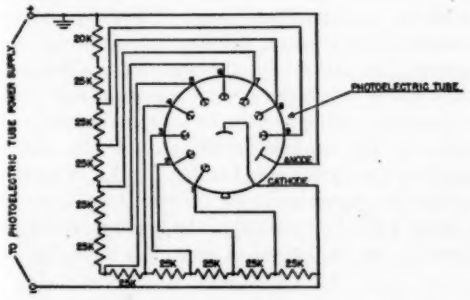

FIG. 9. Wiring diagram for the photoelectric tube.

making of exploratory tests as its response time is less than the recording indicator.

THE PROCEDURE

Although there exists a relatively large number of possible uses for the contrast rendition apparatus model 2, the procedure for its use is substantially the same in all cases. In principle, the measurement of stray light in optical systems is exceedingly simple. Using the apparatus described above, the procedure consists of simply measuring the brightness of the image of the

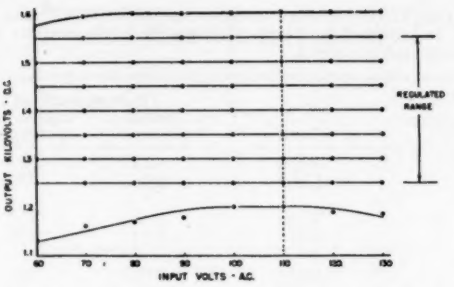

FIG. 10. Output voltage *vs.* input voltage for the photoelectric tube power supply.

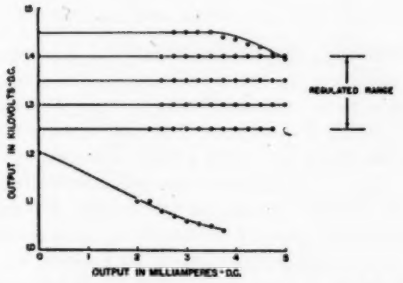

FIG. 11. Output voltage *vs.* output current for the photoelectric tube power supply.

background against which the target is viewed and the brightness of the image of the target. The contrast of the image thus measured is compared on a percentage basis with the inherent contrast of the target which, by definition, is the contrast rendition of the optical system under test.

The definitions of contrast rendition, image contrast, and inherent target contrast, are summarized as shown in Fig. 12. From the "Notes" in Fig. 12, it is seen that the contrast rendition of an optical system is a property of that system and the illumination conditions under which it is used, and as such is a figure of merit of its state of perfection with respect to the amount of non-image-forming light present in the optical system during its use.

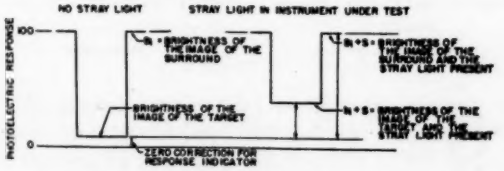
NO STRAY LIGHT
STRAY LIGHT IN INSTRUMENT UNDER TEST
PHOTOELECTRIC RESPONSE
100
0
BRIGHTNESS OF THE IMAGE OF THE SURROUND
BRIGHTNESS OF THE IMAGE OF THE TARGET
ZERO CORRECTION FOR RESPONSE INDICATOR

FIG. 12. Schematic diagram of the type of data obtained in contrast rendition measurements. Notes: 1. The contrast of the image of the target (c_i) is defined to be: $c_i=(B_i-b_i)/B_i$; 2. The contrast of the image of the target is related to the inherent brightness of the target (b_0), the inherent brightness of the surround (B_0), the light transmission factor of the instrument under test (T), and the scattering coefficient (K) by the equation: $c_i=Tc_0/T+K$ where $K=S/B_0$ and $T=B_i/B_0$; 3. Contrast Rendition (C.R.)$=c_i/c_0\times 100$.

TABLE I. Axial contrast rendition for four 7×50×7° binoculars with and without the beam-splitting plate in the contrast rendition apparatus model 2.

Binocular no.	Barrel	Contrast rendition	
		With beam-splitting plate	Without beam-splitting plate
1	Right	92.7	90.5
	Left	91.6	92.7
2	Right	91.6	91.1
	Left	92.2	90.5
3	Right	91.6	90.5
	Left	91.0	90.5
4	Right	91.6	90.5
	Left	89.9	91.1

Because the contrast rendition apparatus model 2 is a new design of an optical testing device, special tests have been made to examine its properties. This examination consisted of two parts. The first involved a study of possible ill effects which might be caused by the use of optical parts in the apparatus and included considerations of a possible procedure for making allowances for such ill effects, should it be necessary. The second part of the examination consisted of an exploration of the various types of tests that could be made using the contrast rendition apparatus model 2.

Validation and Calibration

The validation and calibration of the contrast rendition apparatus model 2 consisted of two parts. The first was concerned with any possible ill effects which may be caused by the use of optical components in the apparatus. The second part deals with the development of a method of making allowances for any such ill effects.

The results obtained by means of the contrast rendition apparatus model 2 were validated by comparing the contrast rendition values obtained with it with those obtained using the earlier model of the apparatus. The results of the comparison indicate that there is practically no difference between the values obtained by the two designs.

Because of the anticipated need for introducing a more complex optical system than exists in the present design of the contrast rendition apparatus model 2, a general method was developed whereby allowances can be made for non-image-forming light introduced by the test equipment itself in stray light measurements. The method consists simply of determining the contrast rendition of an optical system by means of test equipment which has no optical components and comparing this value with the contrast rendition obtained for the test fixture of uncertain properties. By means of such data, the contrast rendition of the test fixture itself can be calculated and the general concepts of contrast rendition applied. This means that only a constant multiplying factor is required to increase the values of contrast rendition measured by the imperfect test equipment to those values that would have been obtained had the test equipment been free from imperfections.

As a test of the above mentioned method of making allowances for imperfect contrast rendition of the test equipment itself, the contrast rendition apparatus model 2 was studied with respect to possible changes in its contrast rendition associated with the presence of the beam-splitting plate, (without the artificial sun in operation). In this study, the contrast rendition of a prism erecting telescopic system was measured for different angular settings of the beam-splitting plate. The changes in the contrast rendition for the different settings are shown in Table I and in Fig. 13. It is evident that there is a variation in the contrast rendition values associated with the rotating of the beam-splitting plate but from a practical point of view, this is negligible.

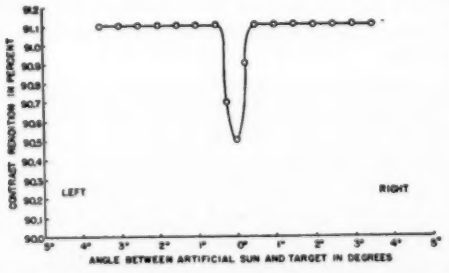

FIG. 13. Contrast rendition for various positions of the beam-splitting plate corresponding to angular separation of the artificial sun and the target using a 7×50×7° binocular.

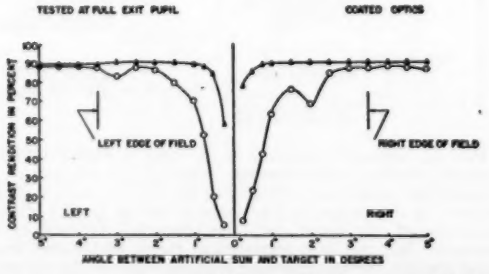

FIG. 14. Axial contrast rendition vs. angle between artificial sun and target for the right barrel of a 7×50×7° binocular.

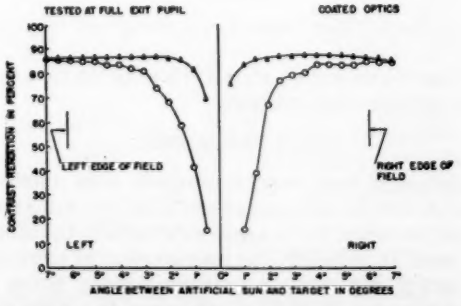

FIG. 15. Axial contrast rendition vs. angle between artificial sun and target for a 5×35×13° lens erecting telescopic system.

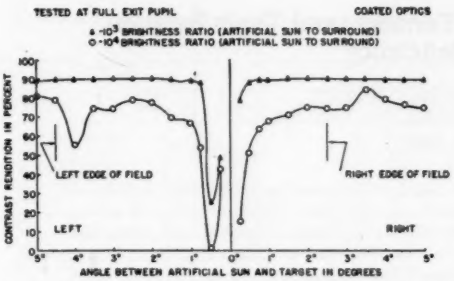
TESTED AT FULL EXIT PUPIL
COATED OPTICS
Δ -10³ BRIGHTNESS RATIO (ARTIFICIAL SUN TO SURROUND)
O -10⁴ BRIGHTNESS RATIO (ARTIFICIAL SUN TO SURROUND)
CONTRAST RENDITION IN PERCENT
LEFT EDGE OF FIELD
RIGHT EDGE OF FIELD
LEFT
RIGHT
ANGLE BETWEEN ARTIFICIAL SUN AND TARGET IN DEGREES

FIG. 16. Contrast rendition *vs.* angle between artificial sun and target for the right barrel of a 7×50×7° binocular rotated off axis 1° to the left of the target.

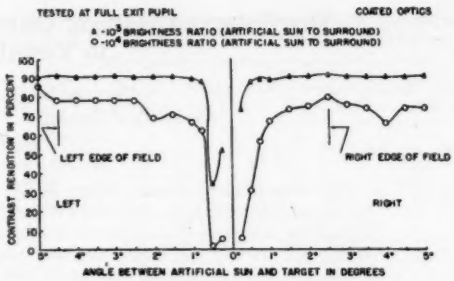
TESTED AT FULL EXIT PUPIL
COATED OPTICS
Δ -10³ BRIGHTNESS RATIO (ARTIFICIAL SUN TO SURROUND)
O -10⁴ BRIGHTNESS RATIO (ARTIFICIAL SUN TO SURROUND)
CONTRAST RENDITION IN PERCENT
LEFT EDGE OF FIELD
RIGHT EDGE OF FIELD
LEFT
RIGHT
ANGLE BETWEEN ARTIFICIAL SUN AND TARGET IN DEGREES

FIG. 17. Contrast rendition *vs.* angle between artificial sun and target for the left barrel of a 7×50×7° binocular rotated off axis 1° to the left of the target.

Experimental Variations

There are two general experimental variations which may be used in the making of contrast rendition measurements of optical systems. The first consists of making measurements with the target aligned on the extension of the optical axis of the system under test and the second with the optical system under test rotated in some off-axis position. For each of these variations, either a uniformly illuminated surround may be used or a non-uniform illuminated surround may be used (such as would result in the use of the artificial sun).

For purposes of illustrating the above mentioned experimental variations, five optical systems were studied. Four were 7×50×7° binoculars and one was a 5×35×13° lens-erecting telescope. In these measurements two different brightness ratios between the artificial sun and the surround were used. The brightness ratios were adjusted by varying the brightness of the surround. In the sample contrast rendition measurements presented, the brightness of the artificial sun was one thousand times as bright as the surround in one case and ten thousand in the other.

THE DATA

The data obtained for the optical systems mentioned above are shown in Table I and in Figs. 14 to 18 inclusive. These data are substantially in agreement with data obtained under similar circumstances using the earlier model of the contrast rendition apparatus, and hence may be regarded as valid values.

DISCUSSION AND CONCLUSIONS

The method of measuring the contrast rendition of telescopic systems in collimated light discussed in this paper has been found to be simple, rapid, and convenient to use. The method described closely simulates field conditions existing in the general use of telescopic

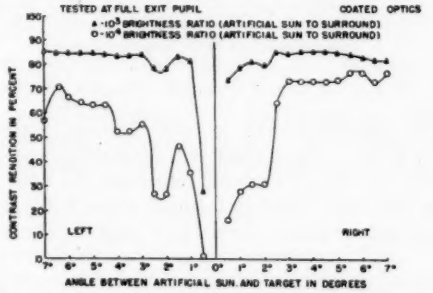
TESTED AT FULL EXIT PUPIL
COATED OPTICS
Δ -10³ BRIGHTNESS RATIO (ARTIFICIAL SUN TO SURROUND)
O -10⁴ BRIGHTNESS RATIO (ARTIFICIAL SUN TO SURROUND)
CONTRAST RENDITION IN PERCENT
LEFT
RIGHT
ANGLE BETWEEN ARTIFICIAL SUN AND TARGET IN DEGREES

FIG. 18. Contrast rendition *vs.* angle between artificial sun and target for the left barrel of a 7×50×7° binocular rotated off axis 1° to the left of the target.

systems in which non-image-forming light introduced into such systems may be the principal factor limiting vision.

The fact that the general concepts of contrast rendition can be applied to contrast rendition equipment itself, results in a much greater freedom in the design of such testing equipment.

The apparatus described in this paper, referred to as "the contrast rendition apparatus model 2" has been found to yield practically the same results for the same telescopic systems studied with an earlier model which did not use optical parts. This indicates that even though corrections could be made for the possible ill effects caused by the use of optical parts required to reduce its dimensions to a practicable size, such corrections are unnecessary for the design described.

ACKNOWLEDGMENT

The authors wish to acknowledge the interest and cooperation received from Lt. Commander C. F. Vance, of the Bureau of Ordnance of the Navy Department, in connection with development of the method of measuring contrast rendition described in this paper.

The Colors of Natural Objects and Terrains, and Their Relation to Visual Color Deficiency*

CHARLES D. HENDLEY** AND SELIG HECHT***
Laboratory of Biophysics, Columbia University, New York, New York
(Received July 20, 1949)

The colors of many natural objects and terrains were determined by visual matching with samples of known spectral reflectance (Munsell's *A Color Notation*). There are only three important groups of colors. Foliage lies in a yellow-green region of dominant wave-length 557–574 mμ. Earths range between yellow and orange-red, 576–589 mμ. Water, sky, and distant objects are blue, 459–486 mμ. Colors of autumn foliage spread into the yellows and reds to the end of the spectrum. The outstanding characteristic of these terrain colors is their lack of saturation, most having an excitation purity of 40 percent or less. Furthermore, objects quickly lose saturation with increasing distance from the observer, so that even in the clearest weather encountered no object over 3½ miles away had an excitation purity of over 11 percent. To the color-deficient observer these colors appear even less saturated than to the normal. In addition, the yellow-to-green region of the spectrum, where the majority of discriminations must be made, is one in which color blinds show poor hue discrimination. Thus, in both hue and saturation the color deficient is at a decided disadvantage in discriminating terrains.

TO determine how a color deficient person will differ from normal in observation under field conditions, it is of course important to know what colors are to be found in terrains, signal lights, colored maps, and in other places where critical judgments of color must be made. At the time this study[1] was made, the necessary data on natural terrains were not available; since then a very comprehensive survey of the colors of soils has been published, including many measurements of foliage.[2] Our results, obtained in a somewhat different way, are presented for comparison with the above survey.

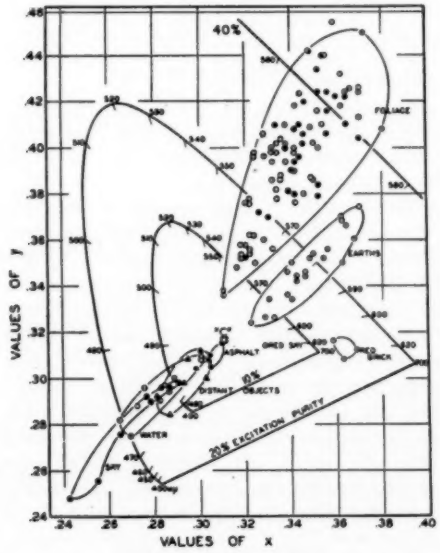

FIG. 1. Colors of natural objects and terrains, summer survey. The central part of the ICI chromaticity chart is shown. The heavy lines mark excitation purities of 10, 20, and 40 percent. Trees, earth, brick, and water are shown by open circles; grass, sky, and asphalt by filled circles. Distant objects are shown by triangles. Some of the points represent single samples, others are averages of several samples.

METHODS

The method used was that of making visual comparisons of natural objects with colored paper samples of known chromaticity. The abridged edition of Munsell's *A Color Notation*,[3] containing 40 hues, was used. Each hue is represented by a page of color chips of varying brightness and saturation.

The color chips were covered by a gray cardboard mask so that only one was exposed at a time, and the objects with which they were compared were viewed through a similar mask exposing the same area on the retina (about 1.4° by 1.4° except in a few measurements of distant terrains, where the area was about 0.7° by 1°). Even this smaller area represents a large section of a distant terrain. At 3½ miles, it is nearly 9000 sq. yd. Such an area is not completely homogeneous in color, and it is an average color that is compared with the standard. Small areas of unusual chromaticity cannot be measured.

The illumination on the objects viewed and on the color chip was approximately the same and was usually mixed sunlight and skylight, having a color temperature of 5000 to 7000°K. Some measurements were made with completely overcast sky, and during rain. No exceptionally clear weather was encountered, and most of the work was done under conditions of light haze.

* This work was done under a contract, recommended by the Committee on Medical Research, between the Office of Scientific Research and Development and Columbia University.
** Now at Department of Physiology, The Ohio State University, Columbus, Ohio.
*** Died September 18, 1947.

[1] C. D. Hendley and S. Hecht, Report No. 269 (Committee on Medical Research, 1944).
[2] Nickerson, Kelly, and Stultz, J. Opt. Soc. Am. 35, 297 (1945).
[3] A. H. Munsell, *A Color Notation* (The Munsell Color Company, Baltimore, 1941), ninth edition.

TABLE I. Colors of natural objects, summer survey, in the Munsell notation.

Object	No. of samples	Distance	Average Munsell color Hue	Value	Chroma
I. Trees and shrubs					
Yellowed tree	1	local	8.0 Y	5	4
Yellowed privet	3	local	9.7 Y	7.7	8
Mulberry	1	local	2.5 GY	5	4
Maple	5	local	3.2 GY	5	4
Wild cherry	4	local	3.3 GY	4	5
Elm	6	local	4.5 GY	5	3.3
Lilac	2	local	5.0 GY	3.5	3.5
Hemlock	2	local	5.0 GY	4	4
Sycamore	7	local	5.0 GY	4.5	4.6
Catalpa	1	local	6.0 GY	4	6
Unclassified trees	6	local	6.0 GY	3.3	2.9
Ash	1	local	6.0 GY	4	4.5
Privet	7	local	6.1 GY	4.5	4.4
Pine	1	local	6.5 GY	5	3
Red oak	2	local	6.5 GY	3.8	4
Ivy	6	local	6.5 GY	5.2	3.6
Willow	4	local	6.9 GY	6	2.8
II. Mixed trees					
Softwoods	5	0.5 to 2 miles	4.6 GY	3.8	3.0
Hardwoods	16	0.1 to 1 miles	5.9 GY	4.2	3.1
Hardwoods	6	1.5 to 4 miles	7.9 GY	6.1	1.1
Mixed woods	2	1 and 2 miles	8.7 GY	3.0	1.5
III. Grasses					
Beach grass	2	local	0.5 GY	6	3.5
Sparse lawn	1	local	2.0 GY	4	3.5
Lawn	31	local	5.5 GY	6.1	4.7
IV. Earths, etc.					
Loam	5	local	8 YR	3.5	0.7
Rich, dark loam	1	local	10 YR	2	1
Dead leaves	2	local	1 Y	5	3
Clayey loam	8	local	2 Y	7	3
Sand	5	local	2 Y	7	2
Sandy loam	4	local	2.5 Y	5	2
Concrete	1	local	2.5 Y	5	1.5
Grass and dead leaves	1	local	5 Y	6	2
V. Water	19	local to 3 miles	1.6 PB	6.3	1.3
VI. Blue sky	11	infinity	2.7 PB	6.2	3.6
VII. Distant objects	11	2 or more miles	2.2 PB	6.2	1.0
VIII. Miscellaneous					
Red brick	3	local	3 R	5	3
Lichen	1	local	7.5 G	4	1
Breaking wave	1	local	10 G	5	1.5
Asphalt	3	local	7.5 PB	5	0.5
Red clouds	1	local	10 RP	8	3

TABLE II. Colors of foliage, autumn survey, in the Munsell notation. Almost all the observations were made at short range.

Object	No. of samples	Range of colors found—Munsell system Hue	Value	Chroma
I. Trees and shrubs				
Sumac	6	4.5 to 7.5 R	3 to 3.5	9 to 10
Unclassified tree	2	4.0 R to 7.5 YR	4.5	4.0
Dogwood	2	6.0 R	4.0	4.0
White oak	9	7.5 R to 4.0 GY	3 to 4.5	3.5 to 8
Butternut	2	2.5 YR to 1.0 Y	3.5 to 5.5	4 to 5
Ash	4	4.0 YR to 5 Y	3.0 to 6.0	4 to 8
Elm	25	4.0 YR to 2 GY	4.0 to 7.0	2 to 10
Hickory	2	8.0 YR	5.5 to 6.0	7.5
Red oak	1	1.5 Y	5.0	5.0
Unclassified	2	6.0 Y	6.0	4.0 to 4.5
Sycamore	5	7.5 Y to 2.5 GY	5.0 to 6.0	3.0 to 4.0
Willow	2	10 Y to 4.0 GY	5.0 to 6.0	2.0 to 3.0
Unclassified tree	1	3.5 GY	5.0	3.5
Unclassified tree	2	6.0 to 7.5 GY	4.0	2.0
Privet	1	7.5 GY	4.5	3.5
II. Lawns	11	2.5 to 8.0 GY	3.0 to 7.0	2.5 to 5.0

Measurements were made at various times of the day between dawn and dusk; most were made near the middle of the day.

The objects were at various distances ranging from 2 feet to 20 miles, and were viewed horizontally, vertically, and at intermediate angles. (No observations were made from aircraft, however.) All measurements were made in New York state, at various locations from the ocean to the St. Lawrence River. About 1000 measurements were made of 235 different objects.

The major part of the survey was made during the summer of 1943 and a supplementary survey was made that autumn.

One observer made all of the measurements. Some of his results were checked by two other normal observers, and no significant differences were found. Most judgments were made under conditions approximating "illuminant *C*." It was found that a moderate change in color temperature had little influence on the judgments of color, provided that object and color chip were under the same conditions of illumination. Therefore, in presenting the data, "illuminant *C*" has been chosen as a reference point. For locating any color, use was made of the trichromatic coefficients of the Munsell colors under "illuminant *C*."[4,5] The values were plotted on large scale coordinate paper, and intermediate colors estimated by linear interpolation.

RESULTS

The chief results are summarized in Fig. 1, which shows the chromaticities of the objects and terrains studied in the summer survey. Only the central part of the ICI color triangle is shown, since most of the samples had an excitation purity of 40 percent or less. In Table I the objects measured are classified into groups of approximately equal hue, and the average color for the group indicated in the Munsell notation. The approximate viewing distance is also shown.

It is striking that only a small part of the region of possible colors is occupied by natural objects. There are only three important groups of colors. Green plants fall in a yellow-green region varying from 550 mμ to 575 mμ in dominant wave-length. Earths and dried vegetation are yellow to orange-red (576 mμ to 589 mμ). Water, sky, and distant objects are blue (459 mμ to 486 mμ). Very few colors were found outside of these groups.

Figure 2 shows the results of the autumn survey, which presents a more varied picture because of the yellow and red leaves of deciduous trees. The Munsell notations are given in Table II. Vegetation covers not merely the summer range of green plants, but also that of earths, and extends beyond in the red to the end of the spectrum.

We have compared these results with the more extensive measurements of Nickerson, Kelly, and Stultz.[2] Their results cover the range reported here for foliage and earths, and in addition they found more saturated foliage colors, as well as much more saturated and redder earth colors. This difference in results may be

[4] Kelly, Gibson, and Nickerson, J. Opt. Soc. Am. **33**, 355 (1943).
[5] Granville, Nickerson, and Foss, J. Opt. Soc. Am. **33**, 376 (1943).

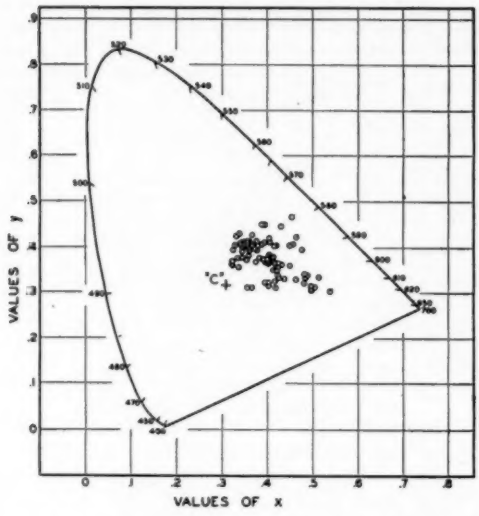

FIG. 2. Colors of foliage, autumn survey. The complete ICI chromaticity chart is shown. Most of the samples are deciduous trees. Their Munsell notations are given in Table II.

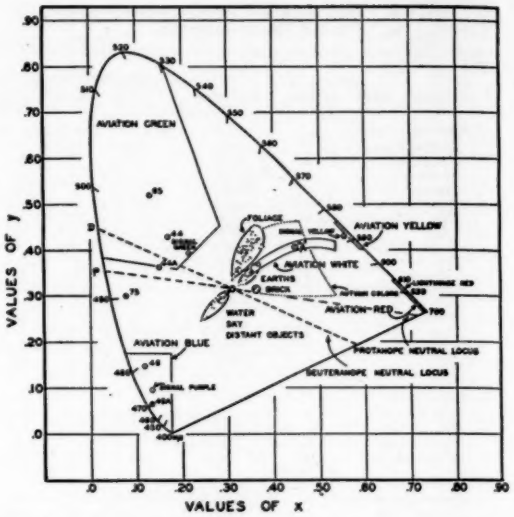

FIG. 3. Terrain colors with illuminant *C* (dots), and colors of signal classes and gelatin filters with illuminant *A* (open circles). The colors are shown on the ICI chart. Specifications of aviation signal colors are shown by solid lines. The neutral loci of the average protanope and deuteranope, in illuminant *C*, are shown by dashed lines.

explained by the fact that we did not make a thorough search for brilliantly colored earths, and also it must be remembered that most of our observations were of terrains seen from a distance.

The influence of atmospheric scattering on the colors of distant objects has been discussed by Middleton.[6] He calculated expected changes in color with increasing distance of observation. The hue which all distant objects approached (about 480 mμ) was very close to the average of about 475 mμ found in this survey.

RELATION TO COLOR DEFICIENCY

The outstanding characteristic of these terrain colors is their lack of saturation; only a few leaves, viewed from close up, had an excitation purity of more than 40 percent. Furthermore, objects quickly lose saturation with increasing distance of observation, so that even in the clearest weather encountered no terrain over 3½ miles away had an excitation purity of over 11 percent. In hazy weather, all characteristic color may be lost when an object is viewed from as little as ¼ mile.

This lack of saturation is of importance in relation to the recognition of terrains by color deficient individuals. Under the term "color deficient" we include all forms of red-green confusers, that is, those who are protanopic, protanomalous, deuteranopic, or deuteranomalous. About 8 percent of all males fall into one or another of these categories. We have excluded tritanopia and tritanomaly from consideration, since the occurrence of persons in these categories is apparently very rare.

Colors appear less saturated to the color deficient eye than to the normal.[7–9] Therefore, since terrain colors are unsaturated to begin with, they will appear even less saturated to the color deficient than to normal vision, and hue in general will tend to be washed out for them.

The degree of difficulty will, of course, depend on the type of defect. The dichromat distinguishes colors in the yellow-green to red region of the spectrum on the basis of saturation differences. The protanope has about 6 iso-color zones in this region, while the deuteranope has about 10.[10,11] Protanomalous observers seem to be very variable in their ability to detect saturation differences. Some may be nearly up to normal, while others have no saturation discrimination at all between 520 mμ and 580 mμ.[12] The deuteranomalous observer is normal or nearly so in the yellow (where normals are weakest in saturation discrimination) but is poorer at other wave-lengths.[13]

In hue discrimination, the critical region for discriminating terrains is 560 mμ–590 mμ. This is a region

[6] W. E. K. Middleton, *Visibility in Meteorology* (The University of Toronto Press, Toronto, 1941), second edition.

[7] J. von Kries, Zeits. f. Sinnesphysiol. **50**, Part II, 137 (1919).
[8] W. Dieter, Zeits. f. Sinnesphysiol. **58**, 73 (1927).
[9] S. Hecht and S. Shlaer, J. Gen. Physiol. **20**, 57 (1936).
[10] F. H. G. Pitt, Report of the Committee on the Physiology of Vision: XIV, Sp. Rep. Ser., No. 200 (Medical Research Council, London, 1935).
[11] D. B. Judd, J. Opt. Soc. Am. **35**, 199 (1945).
[12] W. M. McKeown and W. D. Wright, Proc. Phys. Soc. **52**, 464 (1940).
[13] J. H. Nelson, Proc. Phys. Soc. **50**, 661 (1938).

where the dichromat has no hue discrimination at all. The protanomalous has a hue discrimination curve with a minimum j. n. d. at about 590 mμ–600 mμ, as in the normal, but the curve is raised by a factor of at least 3 in this region.[13] The deuteranomalous wave-length discrimination curve is shifted to the right as well as raised, so that there is a maximum j. n. d. in wavelength between about 570 mμ and 590 mμ. So just in the most critical region for making field judgments, at the dividing line between earths and foliage (575 mμ), the deuteranomalous observer has very poor wavelength discrimination.

Because of the loss of saturation of distant objects, even the normal observer relies on brightness contrast for the principal cue in distinguishing objects at large distances. It is precisely when brightness cues fail that the color deficient becomes severely handicapped, and this shows itself in field experiments.[14]

It may also be noted that dichromats have higher cone thresholds than normals.[15] Their luminosity losses in the blue are insignificant, but at 575 mμ they need about twice as much light as the normal to reach the threshold. So we may predict that at twilight the dichromat will lose his foveal vision a little earlier than normal.

[14] Wallace, Hexter, and Hecht, Research Bull. Hq. 43-6 (Army Air Forces Training Command, October 14, 1943).
[15] S. Hecht and Y. Hsia, J. Gen. Physiol. 31, 141 (1947).

SIGNAL GLASSES AND COLOR FILTERS

As a matter of interest, the chromaticities of a number of representative colored signal glasses with "illuminant *A*" are shown in Fig. 3. The Corning glass filters lighthouse red, signal yellow, signal green, and signal purple are indicated. In addition, the numbered circles refer to Wratten gelatin filters that give highly saturated colors.[16]

The light solid lines indicate the limits specified for aeronautical lights and lighting equipment, the colors being aviation red, yellow, white, green, and blue.

It has long been established that the color deficient tends to confuse green and blue-green lights with white, and red lights with amber or yellow ones. The reason for the first of these confusions is that dichromats have a neutral point in the blue-green. The second confusion results from the dichromat's complete lack of hue discrimination in the yellow and red region. In Fig. 3, the dashed lines show the neutral loci of the average protanope and deuteranope.[11] All colors near these lines, such as signal green, are seen as white by dichromats. Anomalous trichromats undoubtedly see these colors as whiter than the normal person does.

In Fig. 3, the terrain colors of Fig. 1 are also shown, so that the small area they occupy may be compared with the complete ICI triangle.

[16] D. L. McAdam, J. Opt. Soc. Am. 32, 247 (1942).

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 39, NUMBER 10 OCTOBER, 1949

A Photoelectric Reflection-Transmission Densitometer

R. L. Huber and A. B. Rakus
E. I. du Pont de Nemours and Company, Inc., Rochester, New York
(Received June 30, 1949)

A Western Electric ERPD Transmission Densitometer has been modified to permit the reading of diffuse reflection density values. The alterations do not interfere in any way with the original function of the instrument. Construction details of the reflection viewer added to the densitometer and the accompanying optical and electrical changes are discussed. Performance data are briefly summarized, including observations of density values in excess of 2.0 for certain glossy photographic papers.

THE performance of the Western Electric RA-1100 type Transmission Densitometer[1,2] is well known throughout the motion picture industry and wherever rapid, accurate diffuse transmission density values of photographic film deposits are desired. When considering instruments of the same precision and performance designed for obtaining reflectance values of photographic silver deposits on paper, there is, at present, no comparable instrument commercially available.

The authors were confronted with the task of designing and building a reflection densitometer which met the requirements for the production control of a wide range of photographic paper products. In reviewing the problem from many angles and in the interest of expediency it appeared that the Western Electric transmission instrument (often referred to as the ERPD Densitometer) could be adapted to fulfill the specified requirements. This paper describes some of the details relative to the construction of a reflection viewer which when added to the original instrument, provides the necessary additional facilities for a versatile reflection-transmission densitometer.

In proceeding with the design of the viewing attachment it was recognized that the optical mode would, by necessity, be the reverse of that generally employed;

[1] J. G. Frayne and G. R. Crane, J. Soc. Mot. Pict. Eng. 35, 184 (1940).
[2] J. G. Frayne, J. Soc. Mot. Pict. Eng. 36, 662 (1941).

i.e., the sample would be illuminated normally and viewed at 45° rather than the converse. A search through available literature failed to disclose any reference to the proposed method being used in any study of photographic reflection density measurements. However, a brief investigation indicated that the results obtained would be equal or sufficiently close to warrant proceeding with the proposed instrument. A second point for consideration was the design of a method whereby one unique value would be obtained, eliminating the effect of surface orientation. Owen and

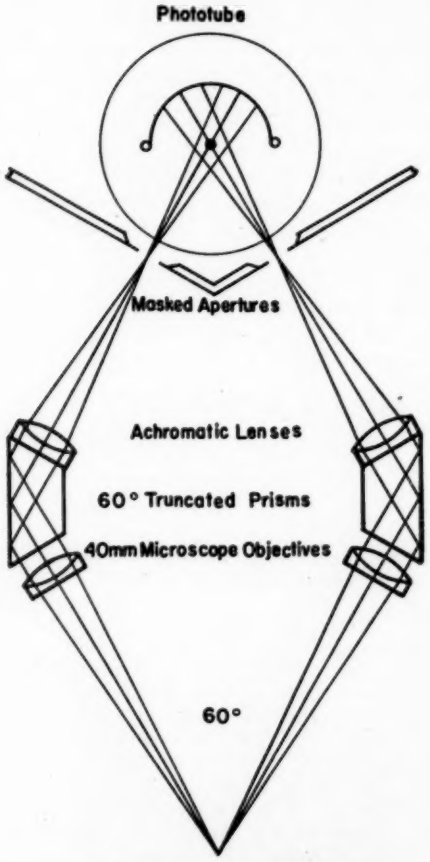

FIG. 1. Schematic of reflection viewer.

Sanders[3] found that a value closely approximating the average can be obtained by measurement of the sample from an angle 45° with the direction of coating. It has also been stated[4] that the orientation differences can be largely resolved by averaging the two density values obtained by single readings taken 90° apart. The authors consider the results more nearly correct when reflected light values are averaged and the corresponding density value obtained. In order to conform to this reasoning the sample should be viewed simultaneously from two directions, 90° apart and both at 45° to the plane of the surface.

[3] R. E. Owen and T. D. Sanders, Phot. J. 73, 398 (1933).
[4] American Standards Association, Sensitometry of Photographic Papers Z38.2.3–1947.

THE REFLECTION VIEWER

The optical system of the reflection viewer is shown schematically in Fig. 1. Reflected light from the sample is picked up by two 40-mm microscope objectives which render the rays essentially parallel. The two paths are brought together by passing each through a 60° truncated prism to which is cemented a lens which focuses the image at an aperture immediately in front of the photo-tube.

It may not be obvious to the casual observer that this system satisfies the viewing requirements for evaluating the reflected light from two directions 90° apart and 45° to the plane of the sample. However, when space relationships are considered. it can be shown that any two surface diagonals from a corner of a cube are 60° apart with respect to one another and that the plane including these diagonals is not 45° with respect to the third surface but approximately 55°. The design of the viewer was materially simplified when it was recognized that 60° prisms could be used.

The viewer (Fig. 2) is built up of brass; the sides are milled from a solid block and the base plate and phototube housing silver soldered thereto to form the main body. The base of the photo-tube compartment is capped with a screw cover obtained from a selsyn motor. The optical components are accessible from the top side of the unit after removing the cover plate. The viewer is firmly mounted to the cast frame of the densitometer optical head by a heavy brass arm fitted with slots for two-directional adjustment. Additional support is offered by the brass tube which carries the leads from the photo-tube to the amplifier.

The microscope objectives are threaded into the walls of the unit and the prisms firmly positioned in a specially designed *C*-clamp. The lenses which are cemented directly to each prism are simple achromats designed to focus the image accurately at the photo-tube apertures. These two apertures are masked such that the phototube response is equal for both optical systems and only light reflected from the same area gets to the photo-tube via the two paths. In addition, all air-glass surfaces in the viewer are low reflectance coated for maximum transmission. The photo-tube is a 929 type, the same as used in the integrating sphere of the basic instrument. The modified densitometer, complete with the reflection viewer, is pictured in Fig. 3.

DENSITOMETER ALTERATIONS

In any reflection densitometer adequate illumination is always a major item for consideration. Although the Western Electric instrument is capable of giving transmission density readings in excess of 4.0, the illumina-

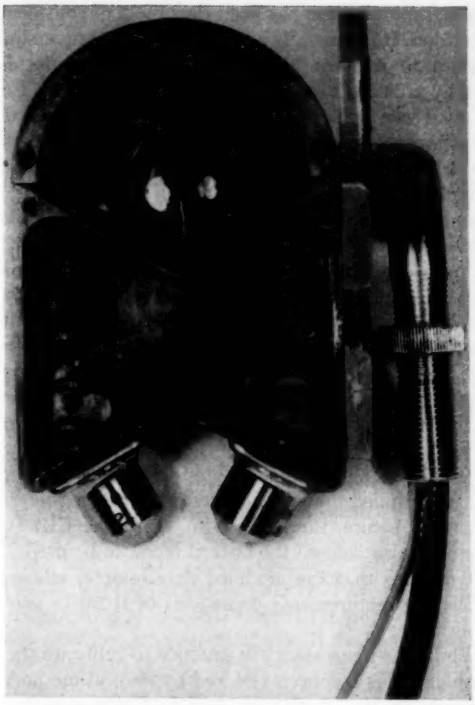

FIG. 2. Reflection viewer showing construction details.

tion level was found to be insufficient to permit using the normal zero-set procedure for reflection measurements on all paper stocks. One method employed to increase the level of illumination was the low reflectance coating of all air-glass surfaces, exclusive of the diffusing block, which gave approximately 50 percent more usable light. Of greater significance was the addition of a second "Visual" filter (Jena OG-1), without the wire screen, in an unused opening in the filter wheel. This change doubled the light flux which, together with the optical coating, accounted for an effective increase of approximately 0.50 density.

The scanning aperture used for both transmission and reflection work normally provides a 2×5-mm illuminated area. For reflection measurements a coated negative cylindrical lens of −4 diopters is inserted in a guide below the 45° mirror, thus enlarging the illuminated area to a 5-mm square. The reflected light collected by the two viewer lenses is masked at the phototube apertures such that a 4-mm square area of the sample is considered. The effective cone angle of each beam is approximately 9°.

The necessity of obtaining the utmost in optical efficiency was also carried over to the electronic circuits. Quite often it is necessary to resort to maximum gain of the amplifier which is accompanied by an increase in noise level and spurious signals. Therefore, it was necessary to shield all circuits more thoroughly. Low loss coaxial cable was used for both photo-tube cathode leads and shielded wire for the a.c. power lines to the separate power supply and voltage regulator, the latter being located about six feet from the densitometer. To minimize spurious signals a good ground is also provided for the instrument.

Since the reflectance values of various paper stocks differ considerably, it is necessary that a more readily accessible gain control be provided than the original "screw-driver-adjust" potentiometer mounted on the amplifier chassis. This control was replaced with a selector switch built up with numerous fixed resistors and mounted on the inclined front wall of the instrument. A second gain control was added on top of the optical head. Here, provision is normally made for a compensating gain increase actuated when the scanning aperture is decreased. Since the same scanning aperture is used for all routine work, the added knob now controls this available gain. Together, the two controls permit a wide selection of gain values equivalent to a density range 1.2 which with the zeroing control on the meter extends the adjustable range to 1.35.

The electronic switching from transmission to reflection or vice versa is handled by a low loss microswitch also mounted on the inclined front wall of the densitometer. This switch is a single pole-double throw type which merely connects the cathode of the required photo-tube to the amplifier.

REFLECTION-TRANSMISSION PERFORMANCE

The alterations and additions made to the densitometer do not interfere in any way with the original function of the basic instrument. In fact, several changes have actually enhanced its usefulness as a transmission instrument. The improved optics, extra electrical shielding and the addition of a cyan filter to the filter selector have extended the use of the densitometer to applications otherwise impossible. The cyan filter, together with the photo-tube response, simulates an orthosensitive receiver such as most dye sensitized paper products.

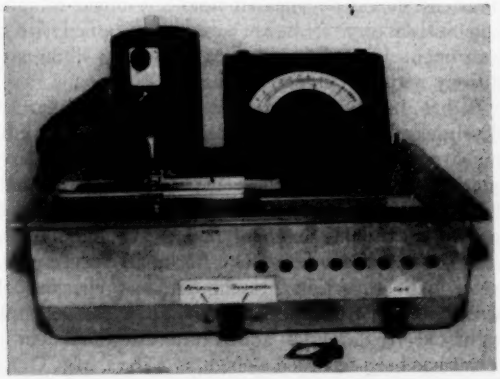

FIG. 3. Completed Reflection-Transmission Densitometer.

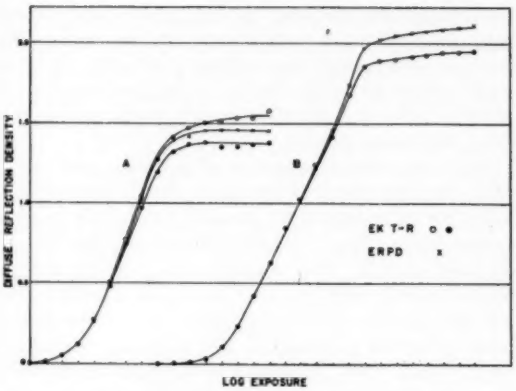

FIG. 4. Plots of typical sensitimetric data obtained with Eastman Transmission-Reflection Densitometer and the modified ERPD Densitometer. (A) Rough textured paper showing single readings taken 90° apart with Eastman *T–R* Densitometer and the corresponding ERPD Densitometer value. (B) Unferrotyped glossy sample exhibiting high maximum density.

The extra transmission of this filter also permits zero-setting of commercial paper products through the base, otherwise impossible with either of the other filters. A feature of the original instrument which has found use for special work is the ability to base-set the densitometer at a meter reading of 1.00 instead of zero. This is accomplished by depressing the 1–2 selector button and adjusting the "zero" control, then depressing the 0–1 selector button to adjust the "calibrate" control for a 1.00 reading. Thus, silver deposits on translucent materials with considerable base density can be conveniently evaluated.

Operation of the instrument for both reflection and transmission densitometry can be carried out under normal illumination levels without having to shield the sample. In fact, a 100-watt Mazda lamp held but a few inches from a paper sample produces no effect whatsoever, even when reading samples at the highest density values. However, light from fluorescent lamps has a noticeable effect on reflection values and the instrument should be shielded from any direct illumination. Satisfactory results are currently obtained with the instrument located in an area indirectly illuminated entirely with fluorescent lighting.

It has been observed that the densitometer will closely check other instruments in the lower and intermediate densities (up to about 1.4 for glossy surfaces) but departs for higher values (see Fig. 4). The maximum density value observed with the instrument for a commercially available unferrotyped glossy photographic paper is 2.20 above base. The Beckman Spectrophotometer with reflection attachment has checked the densitometer within ± 0.01 for all glossy samples evaluated to date. Of added interest, black chiffon velvet has been recorded as high as 2.76 *vs.* $MgCO_3$ and reread with the Beckman as 2.75.

The validity of the higher density values of glossy photographic papers has been the object of considerable attention. It is generally agreed that optical reversibility applies to relative plane surfaces but will not hold strictly true for rough textured materials. From Fig. 4 (A) it is seen that the results obtained with the modified densitometer for a rough surface paper are in good agreement with averages[5] obtained with an Eastman Transmission-Reflection Densitometer which employs 45° illumination and normal viewing. However, when considering glossy surfaces which exhibit little or no azimuth effect, Fig. 4 (B), the two instruments do not agree at the higher density values. Therefore, it may be concluded from these data that the optical mode is not primarily responsible for the observed discrepancies but that they indicate a higher degree of instrument sensitivity or more efficient masking within the viewer of the new densitometer. A discussion of the relative merits of the two methods of illuminating and viewing photographic paper is not within the scope of this paper, nor is it implied that the employed method is preferred. As stated before, the decision to adapt the ERPD instrument also defined the optical mode to be used. The fact remains that the modified densitometer efficiently fulfills the requirements demanded of it in its present application.

While it is unnecessary in practice to calibrate the instrument, this has been checked by several methods in addition to the Beckman Spectrophotometer. When using a glossy fixed-out sample, the alternate blocking of the light entering each lens of the viewer produces an indicated density increase of 0.30. Furthermore, carefully prepared perforated metal plates of density values 0.50 and 1.00 have been inserted in the illumination optical path and have produced the respective increase in indicated density for values in excess of 3.0 with a deviation no greater than 0.01.

The over-all performance of the densitometer, as a research tool and a production control instrument, has been highly satisfactory. In the more than eighteen months of practically continuous use, little maintenance, aside from periodic cleaning and monthly replacement of the light source, has been necessary.

ACKNOWLEDGMENTS

The authors are indebted to several individuals for their contributions to the success of our work, in particular to Dr. D. R. White of the Parlin Laboratory who suggested the possibilities of the Western Electric Densitometer in its present application; to Dr. J. P. Weiss also of the Parlin Laboratory who assisted materially in early experimentations and to Mr. C. R. Keith of the Western Electric Company for his interest and cooperation. To the many others who have assisted in the numerous tests and who have offered their valuable comments, we also wish to express our sincere appreciation.

[5] The term averages used here refers to the values obtained when the two densities are converted to reflectances, averaged and the corresponding density considered.

The Intrinsic Efficiency of the Cathodoluminescence Process*

R. L. LONGINI
Westinghouse Research Laboratories, Westinghouse Electric Corporation, East Pittsburgh, Pennsylvania
(Received January 31, 1949)

It is shown that the efficiency of the cathodoluminescence process is the derivative of the light generated as a function of the energy of the electrons striking the phosphor. Ordinary measurements of light output as a function of electron beam energy are corrected for: electrons not striking the phosphor, loss of electron energy in passing through an aluminum film backing the phosphor, and loss of generated light absorbed in the phosphor layer. The intrinsic efficiency of zinc-cadmium sulfide proves to be much higher than any previously reported "efficiencies."

I. INTRODUCTION

CERTAIN materials emit light when struck by sufficiently high energy electrons. This process of converting cathode-ray energy into radiant energy is called "cathodoluminescence." Phosphors, materials possessing luminescent properties, are very common but relatively few are cathodoluminescent, the majority being excited to emit radiation by absorption of ultraviolet light of greater energy per quantum than the emitted light. Cathodoluminescence is the basis of cathode-ray tubes where an electron beam at several thousand volts and a milliampere or so of current impinges on a phosphor of willemite or "sulfide" resulting in the emission of light. The efficiency of the energy converstion process is sufficient for many practical purposes including projection of television pictures onto moderately large viewing screens.

Cathodoluminescence is not understood in detail. We do know from theory[1] and experiment[2] something of what happens as an electron penetrates matter. The energy of the electron is reduced in many small steps, so that a smoothed curve of the energy as a function of its path length in matter can be expressed by $dE/dx=-m/E$, where E is the electron energy, x is the path length, and m is a constant of proportionality. Now the light emission characteristics produced by cathode-ray bombardment are very similar to, though not identical with,[3] the emission resulting from ultraviolet absorption excitation. It is therefore probable that the direct emission elements of these conversion processes are very much alike and, hence, we are justified in assuming that in cathodoluminescence, as in the case of ultraviolet excitation, electrons somehow are raised to metastable energy levels from which they later fall with the emission of light. The source of energy for this excitation process must be the small steps of energy loss of the primary electrons. Thus the cathodoluminescent process should undoubtedly be considered as a series of individual conversions, each of which consists of the loss of a small amount of energy from the exciting high energy electrons and the subsequent emission of a photon resulting directly from each "step" in the primary electron's energy decay.

This elementary conversion process is the one of interest to physicists, and this paper is a report of an investigation of its efficiency.

We have a simple method of measuring the efficiency of the elementary process. Let us consider the events associated with one electron entering a phosphor crystal with, say, 10 kev energy. It will be slowed in a number of steps, during which period it will excite fluorescence, until it has, for example, only 9 kev energy. The amount of light it will generate after this is equal to that generated by an electron initially with 9 kev. Thus the efficiency between 9 and 10 kev is the difference in the light energy generated divided by 1 kev, the total loss of energy. In the limit the intrinsic efficiency of the elementary process at 10 kev is the derivative dQ/dE taken at 10 kev, where Q is the light generated and E is the electron energy as before.

Though the engineer will still be interested directly in the light output as a function of voltage, we see that the physicist should focus his attention on the derivative. Not only are these fundamentally different, but there may be a factor of two between the mean and intrinsic efficiencies as the light output follows a square law of voltage in some ranges.

Preliminary results of work by Coltman[4] indicate that the mean efficiency of conversion of zinc sulfide phosphor for high energy electrons is in the order of 50 percent. Since the intrinsic efficiency over some region must be higher than the mean we expect that the elementary process efficiency may be high. This high efficiency is of interest for it imposes a stringent criterion on the acceptability of any mechanism proposed for cathodoluminescence and may even influence thought on electron energy loss in matter in general.

Neither the light generated nor the electron energy impinging on the phosphor are directly measurable. Actually we measure the light output of a phosphor and the electron energy in the cathode-ray beam. Before we take the derivative, we must convert this into

* From Ph.D. thesis, University of Pittsburgh Contribution 2-P-49 Physics Department.

[1] N. Bohr, Phil. Mag. **25**, 10 (1913); Phil. Mag. **30**, 581 (1915). H. Bethe, Zeits. f. Physik **76**, 293 (1932).

[2] R. Widdington, Proc. Roy. Soc. **A**, 89 (1914).

[3] Shrader, Lasof, and Leverenz, *Preparation and Characteristics of Solid Luminescent Materials* (John Wiley and Sons, Inc., New York, 1948), edited by Fonda and Seitz, p. 238.

[4] J. W. Coltman, *Preparation and Characteristics of Solid Luminescent Materials* (John Wiley and Sons, Inc., New York, 1948), edited by Fonda and Seitz, p. 417.

light generated and energy of the electrons striking the phosphor.

The absolute output was measured with a photo-cell calibrated by means of an illuminometer and the known spectral distribution of the light emitted by the phosphor. The beam energy was measured with a microammeter and an accurate voltmeter. The problem is

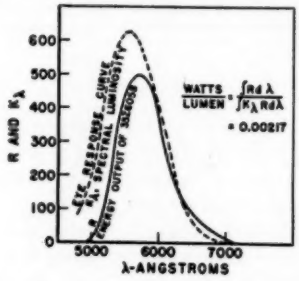

FIG. 1. The spectral luminosity eye response curve, lumens per watt, and the relative power output of our yellow emitting phosphor plotted as functions of wave-length.

then to convert these two sets of measurements into the more fundamental set described, after which the derivative can be taken.

To be certain of the electron energy transferred to the insulating phosphor crystals, the fluorescent screen was backed with aluminum, as is done in the manufacture of modern television tubes. It then becomes necessary to correct for the effects of the aluminum on the electron beam. The aluminum reflects and absorbs some electrons, reduces the energy and scatters others. These effects are all considered. Additional electrons are lost by passing between crystals of phosphor and thus do not produce light. Correction for this requires calculations on the structure of the phosphor layer and determination of the phosphor particle size.

Now the light output is less than the light generated. Some light is lost due to the light scattering power of the phosphor crystals and imperfect reflection of the aluminum. The light loss calculations become very complicated if the position of origin of light is voltage dependent. The situation is much simplified by the choice of large phosphor crystals so that the electrons never have enough energy to pass through them. Thus light is generated in the same crystals at all voltages making the ratio of light generated to that measured a constant. The position of generation of light in a single crystal is relatively unimportant as many internal reflections normally take place before the light escapes from the generating crystal, thus obscuring its point of origin.

This may seem an indirect way of getting the desired results and indeed it is. Ninety percent of the work is involved in making corrections. Many methods of attack were considered and though some of these are simpler in concept the experimental difficulties are so great as to make them impractical with present techniques. However, the calculations performed in this work have the additional virtue of applying to luminescent screens as used, and thus afford a very instructive insight into the behavior of such screens.

II. MEASUREMENT OF LIGHT OUTPUT

W. J. Hushley, of Westinghouse Research Laboratories, measured the light output of a large number of phosphors using a fixed photoelectric pick-up. We will restrict our attention to his results obtained by using a yellow luminescent phosphor made by Radio Corporation of America for television tubes and designated as 33Z605B. This phosphor is Zn—CdS activated with silver and containing, by our analysis, 0.59 Zn and 0.41 Cd on a molar ratio basis. The photo-cell was calibrated for the above phosphor by Coltman, Hushley, and Longini by comparing its output with the reading of a Leeds and Northrup Macbeth Illuminometer.

The spectral emission distribution curve, shown in Fig. 1, was estimated from data given by Shrader, Lasof, and Leverenz[3] and from many typically shaped curves of this type. Because the distribution is fairly narrow and near the position of maximum eye sensitivity, its exact shape is not required and it can be assumed that the 2.17 milliwatts per lumen calculated is not far from correct. (See Fig. 1.)

Some of Hushley's data, converted by means of the above factor to milliwatts output due to one microampere of cathode-ray beam are shown in Fig. 2. The curves represent data on screens of different thickness and various amounts of aluminum backing.

The phosphor crystals were settled onto the glass end of a cup from a dilute potassium silicate solution after which excess liquid was removed and the layer dried and baked. Water was then added to a depth of several millimeters. A drop of collodion solution was allowed to spread over this forming a film, which, after it had become plastic, was brought down to rest as a tent over the phosphor crystals by draining the water from under it. The unit was then dried, after which aluminum was evaporated over the collodion film. The entire cup was then baked in air to 400°C in order to burn out the collodion, leaving the aluminum touching only the high points on the phosphor. The aluminum made contact with a Kovar ring which in turn was connected to the positive terminal of a power supply. The completed cup was then sealed with Apiezon Q wax to an electron gun mounted on a vacuum system and pumped to a good vacuum. The electron gun was energized, the accelerating voltage applied, and the desired readings were taken. The spot size was kept constant with a magnetic field from an external solenoid.

A Radio Corporation of America Type 929 cesium-antimony vacuum photo-cell was used for determining the relative emission for different phosphor surfaces with varying accelerating voltages. The photo-cell was mounted in a light-proof bracket with a window placed against the phosphor in order to make its readings

reproducible. It was found that with this precaution the readings were independent of variations of the spot size and of sideways displacements beyond the limits actually found in the experiment.

A phosphor was operated steadily at 10 kv while a number of illuminometer readings were made of its output. The brightness was measured, and from the known diameter of the spot, the total luminous flux was calculated as 145 millilumens per microampere of electron beam current striking the aluminum-backed phosphor. This corresponds to $145\times 2.17=314$ microwatts per microampere beam current for this phosphor. Under these identical conditions the photo-cell reading was 1.11 microamperes. Thus for phosphors with this spectral distribution the photo-cell read 1 microampere for every 282 ± 20 microwatts output of the phosphor. The probable error indicated is not due to such wide variations in the readings, but involves some possible systematic errors such as errors in the illuminometer, variations of the eyes from normal, non-Lambertian emission, etc.

We now have the measurements of the output and the electron beam energy. If we were to conform to past procedure, we would now divide the output by the input and call this ratio the efficiency. This is as far as anyone has previously gone. The real subject of this paper may be said to start here. We must now correct for electron energy not being transmitted to the phosphor crystals and for light generated that is not emitted by the fluorescent screen due to internal losses. Furthermore we will then take the derivative of the light generated with respect to the voltage of the electrons impinging on the phosphor. The derivative is the incremental efficiency—the efficiency of the individual process of conversion—the quantity of real interest to physicists.

III. CALCULATION OF ELECTRON LOSSES

A. Structure of the Screen

A most important cause of energy loss, and one which has been universally ignored in previous work, is the loss due to electrons passing between phosphor particles without actually striking them. This loss occurs because the crystals of phosphor are random in size and shape and are not formed into a tight mosaic. In order to determine the amount of input energy loss due to this effect, it is necessary to have a knowledge of the geometrical structure of the screen. It is possible to calculate the relevant characteristics of the structure by statistical methods based on a knowledge of the process by which the screen is formed. This aspect of the problem was first attacked by L. P. Hunter (unpublished), but his methods have been revised and extended and further measurements have been made for this paper. The calculations are carried out as follows.

Since the phosphor coating is thin in point of the number of crystals depth, we cannot regard it as uniform. We must, therefore, take into consideration the means of producing such surfaces and determine from the method of production its probable structure.

The coatings are formed by settling particles from a suspension. The particles thus settle in a completely random manner, any interaction entering into their position coordinates of the layer only at the very last when there may be some rolling or sliding to reach positions of lower energy. This will be aided by normal Brownian movement especially in very small particle size phosphors.

We will at first determine the "structure" on the basis of no rolling or sliding. Thus the horizontal coordinates of each particle will be completely independent of those of every other particle. We can, therefore, imagine the particles as settling in any predetermined order without affecting our results. We choose to "settle" the particles in small groups, so that the probability of any two particles in one group overlapping in the final layer is extremely small, i.e., of the second order. The particles in any one group will be randomly positioned above the surface, not clustered. The groups are so selected that the sum of the particle cross sections (projected areas) in any one group is approximately a constant dc. We will define the term "coverage" to be the total cross section of all particles divided by the total area of the layer. Thus,

$$\text{coverage}\equiv C=\frac{1}{S_0}\oint dc, \tag{1}$$

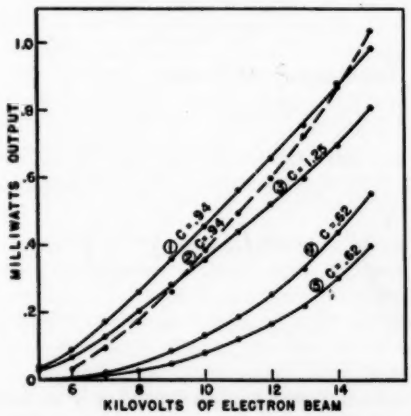

FIG. 2. Measured power associated with light emitted from aluminum-backed zinc-cadmium sulfide phosphor screens at 1 microampere electron beam current for various phosphor screen thicknesses. C is the mean thickness of the layer expressed in mean crystal diameters.

Phosphor screen	1	2	3	4	5
Aluminum thickness (angstroms)	1000	2000	1300	2300	1500
mg phosphor/cm²	6	6	8	4	4
Coverage C	0.94	0.94	1.25	0.62	0.62

where S_0 is the total area. We will say that a particular point of the final layer is covered with r strata if a line normal to the surface at that point passes through r crystals.

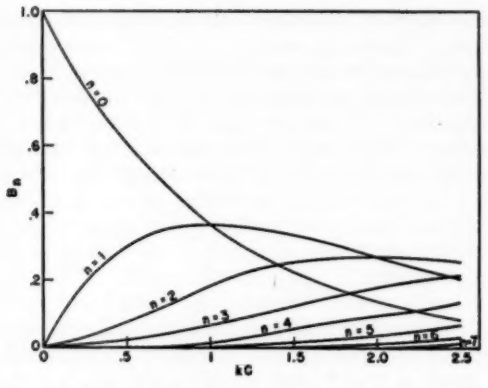

FIG. 3. Probability (B_n) of n particles in a cell as a function of the interaction k and coverage C. Examples: If $k=2$ the cells are twice the mean projected crystal area. If $C=1$ then $kC=2$ and the probability of having the crystal centers of exactly 3 crystals above a cell chosen at random is (from the $n=3$ curve) 0.180. The probability of having one crystal is 0.270 and these cells have only half their area covered.

We now wish to investigate the way in which each group settles. Let S_r be the screen area covered by r *or more* strata and A_r be the *fraction* of screen area covered by *exactly* r strata of crystals. Since S_0 was by definition the total area of the screen, it must therefore also be the area covered by zero or more crystals. Now the probable fraction of dc, the incremental area, that will add to S_r is A_{r-1}, as only crystals added to areas covered by $r-1$ strata will increase S_r. Thus,

$$dS_r = A_{r-1}dc, \tag{2}$$

where, from the definitions above,

$$A_r = (S_r - S_{r+1})/S_0. \tag{3}$$

We can solve the above equations for S_1, and then S_2 and so forth. The general solution is found to be:

$$S_r = S_0\left(1 - e^{-C}\sum_{m=0}^{r-1}\frac{C^m}{m!}\right) \tag{4}$$

and

$$A_r = e^{-C}\frac{C^r}{r!}. \tag{5}$$

The values of S_r and A_r thus depend only on the coverage C and the number of strata r, and are independent of the crystal size distribution.

This is exactly the solution one gets for a slightly different problem. Suppose we divide the total area into small cells each with area g, equal to the mean cross section of the phosphor crystals. If we settle the phosphor on this area as was done above, what is the probability of finding r centers of crystals in a cell chosen at random? The answer is given by the distribution of Poisson which is identical with (5), A_r being the required probability.

This second interpretation of A_r permits an easy extension to the case where phosphor particles have some interaction such as sliding and rolling. The interaction of phosphor crystals is such that there will exist sideways displacements in order to find positions of lower potential energy. The displacement, as measured in mean crystal diameters, will be greater for small than for large diameters as Brownian movement will then enter far more prominently. It should be emphasized, however, that random displacements not resulting in lowering of the potential energy (a type of displacement which might well be associated with Brownian movement) will not essentially alter the previously calculated distribution as the particles of phosphor will still be completely random in position. There must be some forces acting which will create preferred positions. These forces will normally be of very short range, but the range will be extended somewhat by random motions such as Brownian movement, as these motions may move phosphor particles from one potential minimum to a deeper one from which it cannot be moved by similar action.

In general, a phosphor crystal in the surface will have valleys all around it which are potential minima, so that a crystal settling later at this point need not move far laterally to come to rest. Many of the late settling crystals will stay on top of the earlier settling ones, being maintained there by some electrical or frictional forces. Others will settle initially in a position of low potential and, therefore, suffer no lateral displacement.

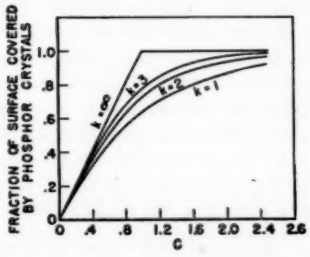

FIG. 4. Fraction of surface covered by phosphor crystals as a function of the coverage C and the interaction k.

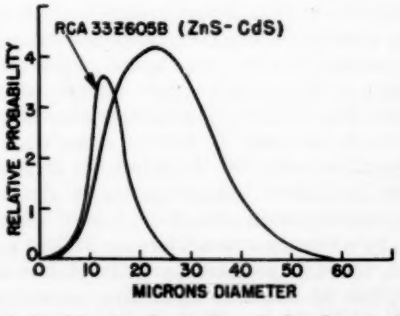

FIG. 5. Particle size distribution of phosphor crystals. The relative probabilities for equal intervals of diameter are shown as a function of particle diameter for the zinc-cadmium sulphide phosphor and for Patterson 1101, a zinc sulphide phosphor used for making optical measurements.

TABLE I. Phosphor particle size and coverage.

Designation of phosphor	Mean crystal height	mg/cm² for $C=1$	Most probable width
Patterson 1101	25 microns	10	22 microns
RCA 33Z605B	14.5 microns	6.4	13.0 microns

We could possibly plot some kind of distribution function showing the probability of lateral displacements as a function of their magnitude. The mean displacement would be small measured in crystal diameters. We could approximate the resultant distribution if we say that the phosphor particles are settled in cells large enough for them to move around in and that, within the boundaries of these cells, they find an absolute minimum of potential energy. These cells would be larger than the mean crystal area by an amount determined by the mean displacement. Suppose that these new cells are k times as big as g, the mean crystal area, then we can recalculate the probability of n crystals having centers above a cell chosen at random by means of the Poisson formula, and find that (5) becomes

$$B_n=\frac{(kC)^n}{n!}e^{-kC}, \tag{6}$$

where kC is the mean value of n over the entire area. Values of B_n are plotted in Fig. 3.

The primary purpose for this calculation has been the determination of how well the electrons are utilized, of the fraction which strike crystals of phosphor. We wish now to determine by calculation the fraction of area uncovered by crystals. All cells with no crystal centers above them are open areas, that is, the fraction $B_0=\exp(-kc)$ are certainly open. But this is not all. If "g" is the mean cross section of all the crystals then the cells each have an area kg. Let us now consider the cells with only one crystal center above them. There are $N_1=B_1/kg$ such cells per unit area that will each have an area kg, and, for N_1 fairly large, their total crystal area will not differ significantly from N_1g so that there will be an open area in these cells equal to

$$N_1gk-N_1g=B_1(k-1)/k. \tag{7}$$

This must be added to B_0 to give fraction open area. If k is greater than two, we find that there are, in general, $N_n=B_n/kg$ cells containing nN_n crystals having a total crystal area nN_ng. The total area of the cells is kN_ng so that the open area is $N_ng(k-n)$ for $k>n$. The fraction of area not covered by crystals is the sum of these terms:

$$\sum_{n=0}^{m} N_ng(k-n)=\sum_{n=0}^{m} B_n\left(1-\frac{n}{k}\right)$$

$$=\sum_{n=0}^{m}\frac{(kc)^n}{n!}e^{-kc}\left(1-\frac{n}{k}\right), \tag{8}$$

where m is the largest integer not greater than k. The curves of Fig. 4 show the fraction of total area covered by crystals, i.e., one minus the fraction of open area which is given by the above equation.

In order to evaluate numerically the above expressions in particular cases it is necessary to assign values to C and k. The actual fluorescent screens are made using a measured weight of phosphor crystals. We must translate this weight into C, the crystal area or coverage, the subject taken up in the next section. The quantity k, which represents the interaction or "smoothing" of the phosphor, can only be estimated roughly but fortunately the quantities dependent on k are relatively insensitive to its value in its expected range. Certainly k is greater than unity, that is, there is necessarily some interaction between crystals, and it seems very unlikely that it is greater than three. In any case an examination of Fig. 4 shows that there will be little error if we assign to k a value of two.

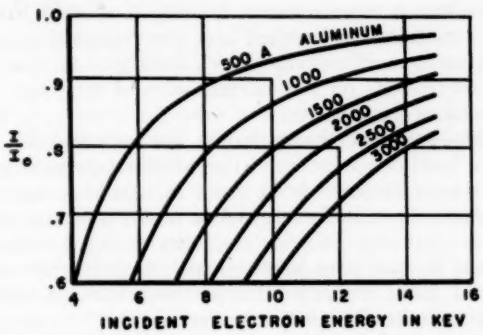

FIG. 6. Calculated ratio of number of electrons transmitted by various aluminum foil thicknesses to the number of incident electrons as a function of the energy of the incident electrons.

B. Determination of Phosphor Particle Size

For the purposes of this work we must know the number of phosphor particles per gram of phosphor and must also know their projected area. If we spread one milligram per square centimeter of a certain phosphor powder on a surface, what is the fraction of the area covered (assuming no overlapping), and what is the distribution of thickness of the phosphor? These questions must be answered in order to calculate the electron penetration and the optical properties of the phosphor coating.

The determination was made by means of photomicrographs using 250 diameter enlargements. From measurements of these pictures (two samples of each phosphor were photographed) a distribution curve was made showing the probability of a certain particle width as a function of these widths. Figure 5 gives the smoothed results of these determinations for the zinc cadmium sulfide phosphor and a zinc sulfide phosphor used for optical measurements. After spending con-

TABLE II. Goudsmit and Saunders' data[a] on observed mean deflection of 900-kev electrons after penetrating aluminum is compared with Eq. (21), which was fitted to it, and with Goudsmit and Saunders' theoretical results.

Centimeters of aluminum	0.0025	0.010	0.025
Observed mean deflection	6°	11°	19°
Calculated from Eq. 21	6°	12°	19°
Calculated by G. and S. theory	6°	14°	23°

[a] S. Goudsmit and J. L. Saunders, Phys. Rev. **57**, 28 (1940).

siderable time on this process, the author feels that much is left to be desired in methods of particle size distribution measurements. Possibly a method based on considerations described later when calculating light scattering may be devised which would give more satisfactory results for the determination of the mean diameter of the particles.

When particles are spread on a microscope slide as was done for the determination of diameters there is a tendency for them to lie down so that they have a minimum potential energy. As a result, we cannot assume that the particles are cubes having a volume equal to the three-halves power of their cross-sectional area. Actually the particles have irregular cross sections so that any "assumption" involving cubical particles would be out of order. Because we want to know the volume of the particles, we have to know the particle depth. This problem was attacked indirectly. The cross-section area of some coarse magnesium oxide particles was measured quite accurately with a projection microscope. The mean value of the area was 3.45×10^{-4} cm² per particle. This value is probably good as the distribution of size was narrow. The weight of 188 crystals was determined as 3.3 mg by direct counting and weighing. From the known density of magnesium oxide of 3.7 g cm^{-3}, the mean crystal volume was determined as being the three-halves power of the base area times $0.78\pm.06$. We have assumed the value 0.78 to be true for our phosphors as well as for the magnesium oxide.

Table I shows the results of measurement and calculation. The mean height is defined by the equation:

$$\text{mean height}=\frac{\int\sigma^3P(\sigma)d\sigma}{\int\sigma^2P(\sigma)d\sigma}(0.78), \tag{9}$$

where σ is the width and $P(\sigma)$ is the relative probability of its occurrence as given in Fig. 5. $C=1$ when the total projected area of all phosphor particles lying in a square centimeter of coating equals one square centimeter.

In preparing the phosphor surfaces the amounts of phosphor used were carefully controlled by weighing. Precautions were taken to make sure that all clusters of crystals were broken up so that we should be able to calculate the value of C very readily from the data just determined on the weight needed for $C=1$. With this we can go back and find the fraction of area not covered by phosphor crystals.

Unfortunately, the electrons are deviated from their path as they go through the aluminum so that they no longer "see" the phosphor thickness along a normal to the surface. If an electron makes an angle ψ with the normal, the phosphor will appear to have a coverage

$$C_e=C/\cos\psi, \tag{10}$$

where C is the coverage as calculated previously and C_e is the coverage as seen by the electron.

Our first step must therefore be to calculate the effects of the aluminum film on the electron paths, as it is the aluminum which we hold responsible for ψ not being zero. In addition, some electron energy will be lost due to the aluminum, and this loss will also be calculated.

C. Electron Penetration of and Scattering by the Aluminum Foil

Experiments on penetration of matter by electrons have been carried out by others, and the results can be expressed approximately[2] by the equations

$$E^2=E_0^2\left(1-\frac{fx}{E_0^2}\right) \tag{11}$$

and[5]

$$I=I_0\left(1-\frac{fx}{E_0^2}\right)fx<E_0^2 \tag{12}$$

where

E_0 = initial electron energy in electron—kilovolts,
E = mean electron energy after going through xg/cm² of material,
f is approximately 5.0×10^5 (kev)² cm²/g, a constant, relative independent of material,
I_0 is the beam current,
I is the transmitted current.

[5] B. F. J. Schonland, Proc. Roy. Soc. **104**, 235 (1923).

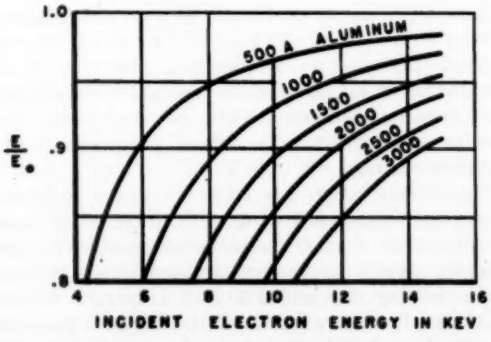

FIG. 7. Calculated ratio of mean energy of electrons transmitted by various aluminum foil thicknesses to the energy of the incident electrons as a function of the latter.

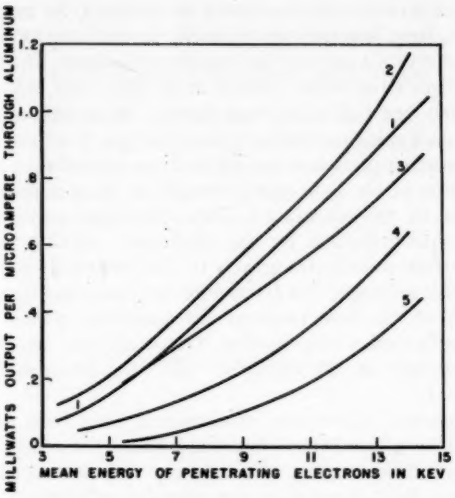

FIG. 8. Output of the various Zn—CdS phosphor screens for one microampere *transmitted* by aluminum, as a function of energy of the electrons after transmission through the aluminum.

Phosphor screen	1	2	3	4	5
Aluminum thickness (angstroms)	1000	2000	1300	2300	1500
Coverage C	0.94	0.94	1.25	0.62	0.62

The current I is reduced primarily by scattering. (Back scattering is referred to as reflection, while side scattering is referred to as absorption.) From these two equations we can easily calculate the effect of the aluminum on transmission of electrons.

The "range" of the electrons is $x_{max}=E_0^2/f$ from Eqs. (11) or (12). For example, with 15-kv electrons the range is 0.45 mg/cm². From Fig. 5 we see that essentially no 15-kv electrons can penetrate the phosphor crystals.

Using Eqs. (11) and (12) we can correct for the beam current reduction and energy loss on traversing the aluminum. The equations are valid only if the electron path length is not much greater than the thickness of material traversed, in other words, only for small scattering. This will be taken up later. Figure 6 shows the fraction of electrons transmitted for beams striking thin aluminum foils. Figure 7 shows the ratios of mean energies of the transmitted to the incident electrons. Using these values, the curves of Fig. 2 have been replotted on the basis of the energy which is available after the electrons have traversed the aluminum.

A relatively simple scattering law can be derived on the basis of classical dynamics. This obviously is not a rigorous procedure when we are considering scattering of electrons, but it gives results in a simple general form that are easily used and fit experimental data rather well.

The scattering processes are assumed completely independent, the energy loss is assumed inversely proportional to the primary electron's energy. A constant of proportionality is chosen to make the calculated and experimental value fit for aluminum. Because the derivation is not rigorous, we will give only the results, which are of engineering value.

The fraction of electrons deflected between ψ and $\psi+d\psi$ is given by

$$W(\psi)d\psi=\sin\psi\exp(-\psi^2/2\varphi^2)d\psi, \tag{13}$$

where B is a normalizing constant for ψ with limits of 0 and π radians, and

$$\varphi=2.2\left(\ln\frac{E_0}{E}\right)^{\frac{1}{2}}\text{ radians for aluminum,} \tag{14}$$

or, in general,

$$\varphi=0.64\left(Z\ln\frac{E_0}{E}\right)^{\frac{1}{2}}\text{ radians} \tag{15}$$

for an element of atomic number Z. These equations fit the data very well, and, therefore, presumably have about the right form.

Equation (14) can be used for determining the coverage as seen by electrons. The electrons have an angular distribution on leaving the aluminum which is given by Eq. (13). The electrons striking the phosphor at an angle with the normal "see" the phosphor as thicker by $(\cos\psi)^{-1}$, that is, as having a coverage $C_e=C/\cos\psi$. We desire to find the mean value of C_e as it enters into

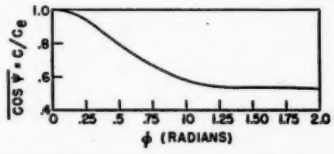

FIG. 9. The mean cosine deviation of path of electrons on penetration of aluminum as a function of the mean deviation φ.

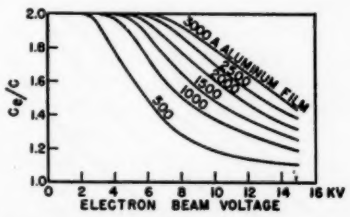

FIG. 10. The ratio of the mean coverage, as seen by scattered electrons, to the actual coverage for various incident energies and aluminum film thicknesses.

electron penetration of the phosphor. This is very complicated, but can be approximated by finding the mean value of $\cos\psi$ from (20) and (21). We have

$$\langle\cos\psi\rangle_{Av}=\frac{\int_0^{\pi/2}\sin\psi\cos\psi\exp(-\psi^2/2\varphi^2)d\psi}{\int_0^{\pi/2}\sin\psi\exp(-\psi^2/2\varphi^2)d\psi}. \tag{16}$$

The limits of the integrals are taken to include only

penetrating electrons. We can then go ahead and calculate $\bar{C}_e$ and from it the fraction of the electrons which pass through holes in the phosphor. Figure 9 plots $\langle\cos\psi\rangle_{Av}$ as a function of φ. These values have been used to calculate $\bar{C}_e$ for various aluminum thicknesses and voltages of the electron beam. The results are plotted in Fig. 10.

We can now make another correction on our light measurement curves—namely, plot the light output for a given current of electrons striking the phosphor. This has been done and the results are given in Fig. 11.

Only one more correction must be made. We have measured the power output, but must know the light generated in order to calculate the efficiency. This requires a determination of the light loss in the phosphor and on reflection by the aluminum.

IV. THE RELATIONSHIP OF LIGHT GENERATED TO LIGHT OUTPUT

The light output of an electron phosphor will be calculated as a function of the light generated and of the various optical properties of the phosphor layer. These properties include the reflectance of the aluminum back-

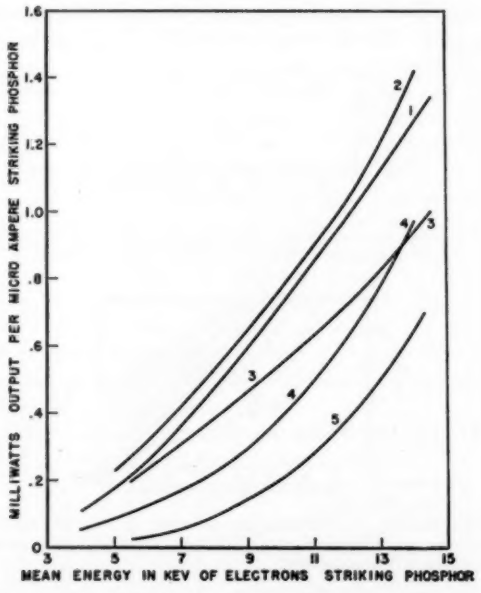

FIG. 11. Output of the various phosphors for one microampere of electrons *striking phosphor* crystals, as a function of energy of electrons after transmission through the aluminum.

Phosphor screen	1	2	3	4	5
Aluminum thickness (angstroms)	1000	2000	1300	2300	1500
Coverage C	0.94	0.94	1.25	0.62	0.62

ing; the scattering and absorbing power of the phosphor crystals; the nature of the glass to phosphor junction and glass to air reflectance.

Before we go into the determination of these individual values, it is better to examine the entire picture to see how each factor enters the problem. As we have seen, light is generated primarily in the layer of phosphor crystals nearest the aluminum backing. The light emerges from these crystals in all directions, approximately one-half going back toward the aluminum and one-half going toward the glass envelope. Since the nonfluorescing phosphor and glass act as reflectors, a good portion of the light going toward the glass is reflected back to the aluminum. Thus the light may suffer multiple reflection by the aluminum backing which therefore contributes greatly to the losses as it is not a perfect reflector. The fraction of loss obviously depends both on the reflectance of the aluminum and of the phosphor-glass combination. Obviously any losses by absorption in the phosphor will also diminish light output.

Specular reflectances of aluminum have been measured by a number of observers.[6] Since the reflectance shows little angular dependence, the reflectance to diffuse light is equal to the specular reflectance. The best aluminized mirrors are on glass and these measure, for good mirrors, anywhere from 0.80 to 0.93 reflectance. Probably 0.8 is a good estimate for the reflectance of the aluminum used as a backing to the phosphor.

We must next determine the multiple reflection between the aluminum and phosphor layer in order to find the loss due to the imperfect aluminum reflections. For this purpose we must evaluate the reflection and transmission factors for the phosphor layer.

Before we take up the case of multiple reflection, we will take up the much simpler case of a perfectly transmitting clear layer of phosphor. In this simple case one-half the generated light will go out directly to the observer for measurement and the other half will be emitted after a single reflection from the aluminum. As a result the total output is

$$\tfrac{1}{2}L+(0.8)\tfrac{1}{2}L=0.9L, \tag{17}$$

where L is the light generated.

Actually, because of the glass, there would be multiple reflections in this case too with somewhat increased loss. Suppose we assume that the glass has mean reflectance of 0.1. Then, of the $0.9L$, only the fraction 0.9 or $0.81L$ will come out directly, while $0.09L$ will be sent back to the aluminum. Of this latter $(0.8)(0.09L)=0.072L$ will be reflected but not all this will get out of phosphor. A detailed analysis shows that the total output would be

$$0.9L(0.9+0.8(1-0.9)+(0.8)^2(1-0.9)^2+\cdots)$$
$$=0.81L+\frac{0.9L(0.08)}{1-0.08}=0.888L. \tag{18}$$

Now it turns out that the phosphor crystal deposit is very reflecting which increases the loss by multiple

[6] J. Strong, *Procedures in Experimental Physics* (Prentice-Hall, Inc., New York, 1942), p. 375. F. Benford, J. Opt. Soc. Am. 32, 174 (1942).

reflection on the aluminum considerably, although the phosphor has little absorbing power in itself.

A. Reflection and Transmission Factors

The differential equations for transmission and reflection have been set up and solved by Coltman *et al.*, Hamaker,[7] and others. These results have been analyzed and extended by Longini.[8] The reflection and transmission factors are found to be

$$R=\frac{\frac{s}{p}\sinh pt}{\cosh pt+\frac{s+w}{p}\sinh pt}, \tag{19}$$

where $p\equiv(w(2s+w))^{\frac{1}{2}}$ for brevity and

$$T=\frac{1}{\cosh pt+\frac{s+w}{p}\sinh pt}. \tag{20}$$

We can see immediately that

$$\frac{R}{T}=\frac{s}{p}\sinh pt. \tag{21}$$

permitting an easy solution for p small, we have Eq. (7) of Longini[8]

$$R=Ch/(1+Ch) \tag{22}$$

and

$$T=1/(1+Ch), \tag{23}$$

where C is the "thickness" expressed in phosphor coverage, and $h=s$ divided by the mean particle diameter.

Coltman *et al.* have pointed out that s and p can be evaluated most accurately by direct experimental determination of R/T, and then varying s and p to fit the data. This is the best means as no absolute reflection standard need be accepted in advance, nor are surface phenomena too important. A series of experiments were made on zinc sulfide phosphor crystals using an Ulbricht photometric sphere and a Leeds and Northrup Macbeth Illuminometer.

The phosphor particles were deposited on a glass slide in various amounts and the ratio of reflected to transmitted light was measured. Corrections were made for the effect of the glass slide by independent measurements. Half of a diffusing standard was covered with glass, the other half left free and the two were compared. The glass has in effect a reflectance of about ten percent.

The quantities h and p were then evaluated by curve fitting. The best values were found to be $h=1.02$ and $p=0.07$ for $t=1$ at 10 mg/cm² of zinc sulfide (Patterson 1101) phosphor. The 10 mg/cm² value was chosen as this is the density of deposit needed to make $C=1$. This method is the most sensitive to variations in the value of p, but here p was so small that there is considerable uncertainty in its value. The coefficient h, however, is very well determined.

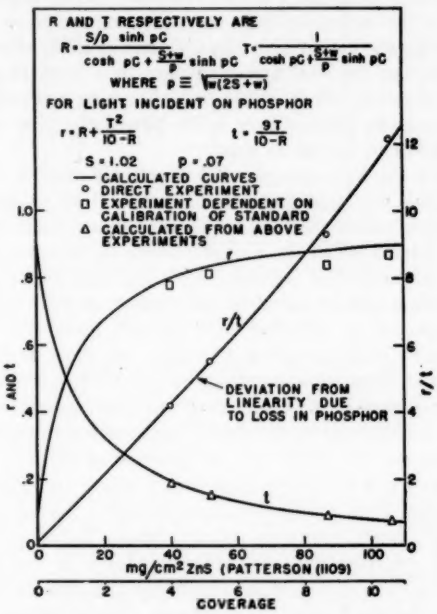

FIG. 12. Reflection (r) and transmission (t) factors for phosphor on glass (glass reflectance 0.10) which involve reflection (R) and transmission (T) factors for phosphor screen alone.

Independent measurements were made of the reflectance of the phosphor deposits. The reflectance was measured by comparison with a calibrated standard furnished with the illuminometer. The calculated reflectances were slightly high compared with the observed values, a discrepancy believed due to an error in the standard calibration.

The calculated curves and experimental points are shown in Fig. 12.

The absorption coefficient w turns out to be 3×10^{-3}, a negligible quantity when one is dealing with very thin phosphor surfaces. For zinc-cadmium sulfide the reflectivity drops slightly in the red,[3] indicating that p has increased to about 0.2. We then have $w=0.02$, still very small.

The experimental points and calculated curves have been plotted in Fig. 12. The curve for r/t fits well within the experimental error of the weight measurements. Using the same values of p and s we find $R_1=0.505$ and $T_1=0.494$ for $C=1$, showing that the loss is negligible.

B. Utilization of Generated Light

We will make use of these values in a very dubious manner. We can justify the means, however, with the

[7] Coltman, Ebbinghausen, and Altar, J. App. Phys. **18**, 530 (1947). H. C. Hamaker, Philips Res. Rep. **2**, 55, (1947).
[8] R. Longini, J. Opt. Soc. Am. **39**, 551 (1949).

statements that the results are not too critical of the exact method employed, and that the approximation is probably pretty good anyway.

In calculating interaction between crystals in settling we assumed the surface was divided into uniform cells equal in area to some constant times the mean crystal projected area. We will now calculate the light getting out of the individual cells as though they obeyed individually the laws derived by Coltman *et al.* In effect we are saying the cells are equivalent to a layer of great lateral extent which is not true. There is, however, an approximate light balance in the lateral direction which turns out to be just as good.

Half the light generated in a cell will start forward and half will start backward. Each will be subject to multiple reflections as it works its way out. That which goes back toward the aluminum can be thought of as multiply reflected between the aluminum and all the phosphor and glass, while that going forward can be thought of as reflected between the glass plus any "excess" non-fluorescing phosphor on the one hand, and the aluminum plus generating phosphor on the other. Figure 13 shows this multiple reflection schematically for the case of the light initially going toward the aluminum. As before n is the number of crystals in a cell and k is the interaction factor which controls the cell area. Thus n/k is the mean value of C, the coverage, for cells containing n phosphor crystals. The sum of the terms representing the outgoing light is $(\frac{1}{2}LR_0t)/(1-R_0r)$ where, as before, r and t are the reflection and transmission factors, respectively, for phosphor on glass. R_0 is the reflectance of the aluminum. This sum must be added to the similarly derived terms for the

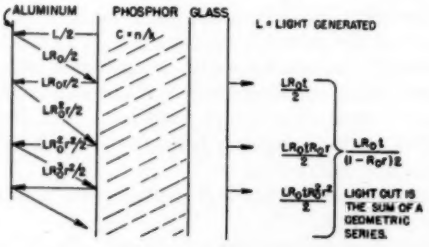

FIG. 13. Schematic arrangement for calculating light utilization. Only light starting toward the aluminum is considered here.

light starting in the forward direction. The sum of these terms has been calculated and is given in Fig. 14.

Areas with no phosphor do not, of course, generate any light, thus no light strikes the glass normal to the surface of these areas. Certainly some light gets out by striking the tube end of cells with no phosphor. This results in a lack of complete balance in the lateral radiations as no light can originate in these cells to be emitted from other cells. The actual fraction of light out will lie somewhere between the transmission calculated on the basis of (1) each cell transmitting the light it generated, (2) all cells transmitting together all the light generated. Fortunately these two values are very close to each other as seen in Table III, so that the mean is certainly close to the correct value.

C. The Elementary Process Efficiency

We can now calculate the true incremental or elementary process efficiency. The slopes are measured in Fig. 11 and these are divided by the above ratios to give the desired results. This has been done and the results are plotted in Fig. 15.

The different surfaces all show efficiencies at 15 kv that are above ten percent. None are as high as 50 percent estimated by Coltman for much higher energy electrons, but the efficiency at 15 kv is still rising with bombarding electron energy. Probably the efficiency goes up to a high value and remains there. This is deduced from the response of the photo-multiplier radiation detector to high energy beta-particles. Calculations of light output of a phosphor fit experiments well on the basis of a constant high efficiency and known electron penetration characteristics. Thus at high energies of the bombarding electrons the "mean" and incremental efficiencies approach equality.

The curves of Fig. 15 seem to indicate that the phosphor efficiencies behave very differently as a function of voltage. This may be due in part to an error in the estimate of the amount of aluminum. Phosphor number 2 is shown with 1300A aluminum as well as the 2000A it is actually believed to have. This changes the shape below 8 kv but does not alter the higher energy efficiencies appreciably.

It is interesting to compare the mean phosphor efficiency at 14 kv to the intrinsic values. Using Fig. 15 we tabulate the results.

The intrinsic or differential efficiency has a fairly large probable error as it is difficult to estimate the derivative when the data itself is not precise. The curves given in

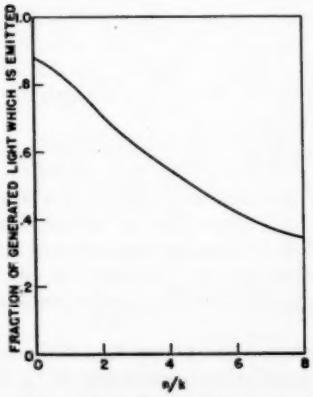

FIG. 14. Fraction of the light generated recovered through the end of a cell with n crystals, interaction factor k, aluminum reflectance 0.8, and effective glass reflectance of 0.1. A lateral balance of light flux is assumed.

Fig. 15 are the result of smoothing after each correction was applied. Because of this lack of precision the curves should not be extrapolated at all. It can be seen from curve 6 of Fig. 15 that the proper determination of aluminum thickness is not very important in the calculation of intrinsic efficiency. Less aluminum means more energy through and therefore one would estimate a decreased efficiency calculated for the same light output. However, the electrons are not scattered as much on passing through the thinner film so that more electrons pass between phosphor particles. The shape of the calculated efficiency curve does, however, depend on the selection of aluminum thickness.

TABLE III. Table of ratios of light emitted to light generated.

Phosphor	Each cell transmitting light it generated	Average transmission including open areas	Mean
1 and 2	0.790	0.803	0.796
3	0.770	0.779	0.774
4 and 5	0.807	0.828	0.818

TABLE IV. Phosphor characteristics at 14 kv, 1 microampere.

Phosphor	Watts output	Efficiency Mean	Efficiency Intrinsic
1	8.7×10^{-4}	6.2%	18%
2	8.7	6.2%	25%
3	7.0	5.0%	14%
4	4.4	3.4%	25%
5	3.1	2.2%	19%

CONCLUSION

The efficiency of the elementary conversion process is up to 25 percent at 15 kev energy of the bombarding electrons and seems to be increasing further. When one considers that the mean wave-length of emitted light is 5700A corresponding to only 2.15 ev energy, the incident electrons must be loosing energy in steps in the order of magnitude 8 ev if only one excitation results per step of energy loss. For light emitted in the blue, assuming the same magnitude energy loss jumps, the 50 percent efficiency estimated by Coltman does not seem at all improbable.

The efficiency reported here is greater than previously reported work for several reasons. Correction for the loss of electron energy in penetrating the aluminum, loss of electrons in passing through holes in the phosphor, and loss of light before measurements are made account for a factor of about 1.5, while taking the derivative to find the efficiency of the elementary process accounts for a further factor of from 2 to 4. These factors depend on the phosphor layer formation as described and may vary greatly from one situation to another.

One might at first think that the phosphors composing the different screens should all have the same intrinsic efficiency. Due to handling, however, impurities are introduced which greatly effect the output; even color changes have been observed in some screens.

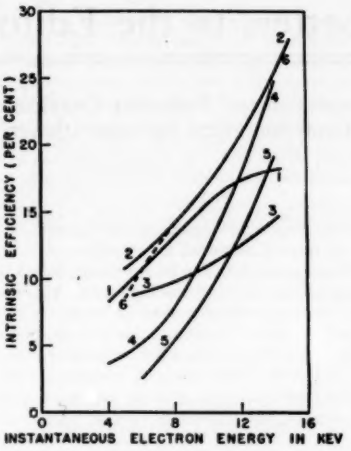

FIG. 15. The efficiency of the individual process of conversion as a function of instantaneous electron energy.

Phosphor screen	1	2	3	4	5
Aluminum thickness (angstroms)	1000	2000	1300	2300	1500
Coverage C	0.94	0.94	1.25	0.62	0.62

Curve 6 is phosphor screen (2) recalculated as though it had an aluminum thickness of only 1300A instead of the 2000A which it presumably has.

ACKNOWLEDGMENT

I wish to express my appreciation to Dr. J. W. Coltman for his considerable help in discussions and for his aid in preparing this manuscript. I also wish to thank the Westinghouse Research Laboratories for the use of their facilities and for permission to use the data taken by Dr. Walter Hushley.

Physics Abstracts

For many years the dues of the American Physical Society have included a subscription to Physics Abstracts, published by the Institute of Electrical Engineers of London. There has been, however, no convenient method by which persons not members of the A.P.S. could receive this publication. The Optical Society of America has now instituted an arrangement under which any O.S.A. member can receive Physics Abstracts at an annual subscription price of $4.50. All correspondence in this connection should be addressed to the American Institute of Physics, 57 East 55th Street, New York 22, New York.

Letters to the Editor

Transmission and Reflection Coefficients of Aluminium Films for Interferometry

M. F. CRAWFORD, W. M. GRAY,* A. L. SCHAWLOW, AND F. M. KELLY**
McLennan Laboratory, University of Toronto, Toronto, Canada
July 22, 1949

THE transmission and reflection coefficients of aluminium films on quartz, prepared for interferometry in the ultraviolet, have been measured. The films were deposited by evaporation in vacuum, using both low and high rates. At the low rates of evaporation the films were deposited in twenty minutes to two hours. At the high rates the evaporation was completed in about twenty seconds. A low rate gave a film with an extremely uniform appearance, but with a low reflection coefficient which made it unsuitable for interferometry. A high rate gave a film with a mottled appearance, but with a much higher reflection coefficient.

The reflection coefficients were measured by the method described by Childs,[1] improved by calibrating the photographic plates for intensity with a stepped-slit continuous spectrum. The stepped-slit was illuminated by the light from a hydrogen lamp diffused by a ground quartz plate. The etalon was placed in its normal position in the spectrograph (following the collimator lens) and a small rectangular aperture was used instead of a slit. The reflecting surfaces were inclined slightly to produce successive images of the aperture in a line perpendicular to the direction of dispersion on the photographic plate. If the intensity of the first image is taken as unity, that of the mth is $I_m=(R^2)^{m-1}$, where R is the geometrical mean of the reflection coefficients of the pair of films. Thus a plot of $\log I_m$ against m is a straight line whose slope gives a value of R with an error of ± 0.3 percent.

The transmission coefficients were determined from stepped-slit spectrograms. Two exposures of equal duration and constant source conditions were made, the first with the film in the path of the light and the second without the film. A calibration exposure was made under the same conditions except that the distance of the hydrogen lamp from the diffusing screen was adjusted so that the density range included the densities of the strong steps on the first exposure and the weak steps on the second. At a given wavelength the intensities of the steps on the first and on the second exposure, determined from their measured densities, were plotted against the nominal intensities of the steps as used for the calibration curve. The points for each of these two exposures lie on a straight line through the origin whose slope can be determined to within 2 percent. The ratio of the slopes of the two lines gives the transmission coefficient of the film for this wave-length.

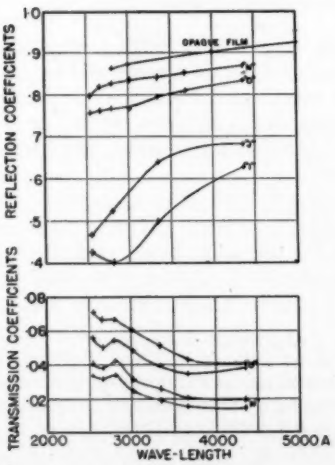

FIG. 1. Reflection and transmission coefficients of aluminium films on quartz.

The results of measurements on typical pairs of films are shown in Fig. 1. The "*N*" and "*O*" films were prepared by rapid evaporation. The transmission coefficients of each film of these pairs are shown, together with the geometrical mean of their reflection coefficients. The "*I*" and "*J*" films were prepared by slow evaporation and only their reflection coefficients are shown. Their transmission coefficients, measured using a photronic cell and a white light source, were 0.062 and 0.04 respectively. The transmission coefficients of the "*N*" and "*O*" films determined in the same manner agreed with the values for blue light obtained by the photographic method. A plot of the reflection coefficient of a good opaque film, taken from a recent paper by Hass and Scott,[2] is included for comparison.

Since the "*J*" and the "*O*" films have approximately the same average transmission coefficients in the visible part of the spectrum it is interesting to compare their reflection coefficients. It will be seen that the reflection coefficients of the "*O*" films is much better for all wave-lengths. The sharpness of the Fabry-Perot fringes also confirmed that for the same transmission coefficient in the visible the films prepared by rapid evaporation have much higher reflection coefficient at all wave-lengths than those prepared by slow evaporation. The marked differences in the properties of the films are thought to be due to a difference in structure determined by the rate of deposition.

The properties of slowly and rapidly deposited films are being correlated with the structures of the films as observed with the electron microscope by G. D. Scott and R. S. Sennett of this laboratory. Their results on silver films show that both the optical properties and the structures are very dependent on the rate of deposition.

* Holder of a fellowship from the Research Council of Ontario.
** Holder of a fellowship from the National Research Council of Canada.
[1] W. H. J. Childs, J. Sci. Inst. 3, 97 (1926).
[2] G. Hass and N. W. Scott, J. Opt. Soc. Am. 39, 179 (1949).

On the Standard Source for Low Level Photometry

J. TERRIEN
Bureau International des Poids et Mesures, Sèvres, Seine-et-Oise, France
July 25, 1949

THE definitions and agreements of photometry are now well-settled and they received the sanction of the Comité International and the Conférence Générale des Poids et Mesures.[1] Although these definitions, the single official ones today, have been adopted without any limitation of the scope of their validity, they are likely to be modified in the field of low luminance levels, in order to take into account the Purkinje effect, and new international agreements, less simple that the present ones, are likely to be adopted in the future. An almost unanimous agreement has already been achieved at the eleventh session of the C.I.E. (Paris 1948); it settles the choice of the primary standard source for all future photometric system and the magnitude of the unit of what will be, in these systems, the analogy of luminance (or photometric brightness). The text of the resolution adopted unanimously by all the nations (except Australia) which attended the C.I.E., and in particular by the U. S. A., reads as follows:

"Light and Vision. Resolution 2. It is recommended that, when for special purposes the luminous effects of radiant energies of various spectral compositions are evaluated by methods that do not rest on the standard luminosity function adopted by the C.I.E. in 1924, the unit of the quantity corresponding to luminance (photometric brightness) should always be chosen so that this quantity has the numerial value 60 (c.g.s. system) for a black bodv at the temperature of freezing platinum."

This recent decision introduces a logical uniformity in the definitions and it is a pity that K. S. Weaver, when he wrote his remarkable paper "A provisional standard observer for low level photometry"[2] seemed to ignore the above resolution. Every table, from II to XI, appearing in his paper, is founded upon the adoption of a standard source at a color temperature of 2360°K; the reasons of this choice are discussed in the text, but no mention appears of the resolution of the C.I.E. supporting the black body at the freezing point of platinum as a standard source. This is the reason why we felt it useful to recall this resolution, which is to be taken into account in every draft of a system of scotopic or mesopic units.

Perhaps is it also useful to mention that when a dark-adapted observer, whose spectral luminosity curve is then certainly not the 1924-C.I.E. international one, is performing equality of luminance settings with a visual photometer head, the results he obtains may not be expressed in terms of the classical luminance units (stilb, for instance); such an observer uses a not yet well-defined system, to be later more precisely fixed thanks to works similar to those of Weaver, Crawford, and others. But as soon as possible, it is largely desirable that the comparison field of the visual photometers used in these conditions should be adjusted at 2042°K, since it is certain that in any future system, the black body at that temperature shall be the reference standard; for this standard and for this one only, we know that the quantity corresponding to luminance will vary proportionately to the energetic luminance. The use of a 2360°K lamp, formerly adopted in several countries, could then be abandoned in order keep in the lines of the C.I.E. recommendations.

[1] Comité International des Poids et Mesures, Procès-Verbaux des séances Deuxième série. Tome XX, 1945–1946, p. 119.
[2] K. S. Weaver, J. Opt. Soc. Am. 39, 278 (1949).

Photosensitivity of Metal Halides as Affected by Adsorbed Ions*

W. A. Weyl and D. P. Enright
Department of Mineral Technology, The Pennsylvania State College, State College, Pennsylvania
July 15, 1949

THROUGH the work of S. Sheppard it is known that the presence of traces of sulfur greatly increases the photosensitivity of silver halides. W. A. Weyl and T. Forland[1] attribute this phenomenon (sensitivity speck) to the asymmetrical force fields caused by the grouping, $Br^- - Ag^+ - S^{2-}$. On the same basis, the greater sensitivity of the "silver-body" as compared with the "bromide-body" was explained.

The surface of a crystal produces an asymmetrical force field which may have a considerable depth action provided the polarizability of the ions is high. The depth action may be described as a broadening of the distribution of binding forces within the crystal and as such is conducive to photolysis. The disturbance created by the surface may be increased or decreased by bringing foreign atoms or molecules onto the crystal surface.

It was found that thallous ions and lead ions strongly increased the photosensitivity of silver chloride. Among the anions, fluorine was found to be particularly efficient as a sensitivity speck for both AgCl and AgBr. Lead bromide when precipitated from a solution containing excess Pb^{2+} ions does not easily darken when exposed to sunlight. However, the same compound as a "bromide-body" is strongly photosensitive.

In a similar fashion the photosensitivity of mercuric iodide can be increased or decreased by bringing foreign ions into its surface layer. This photolysis is of particular interest because it may serve as a trigger reaction to set off a phase transformation. For example, the high temperature form of the mercuric iodide is stabilized by embedding it in a polarizable matrix, such as paraffin. This treatment makes it possible to retain the yellow form of HgI_2 at room temperature. Exposure to sunlight, however, produces photolysis and the resulting centers of perturbation cause the yellow crystals to convert into the red form which is stable below 126°C. The rate of this reaction depends upon the particle size of the mercuric iodide and on the nature of the surface ions.

* Sponsored by the ONR and carried out under Contract N6 onr 269 Task Order 8 NR 032–265.
[1] W. A. Weyl and T. Forland, "Crystal Chemistry and Optical Properties of Defect Structures." ONR Technical Report No. 2 Task Order 11, Contract No. 6 onr 269–11 NR 081–060.

Photosensitivity of Mechanical Mixtures of Non-Sensitive Compounds*

W. A. Weyl and A. I. Cramer
Department of Mineral Technology, The Pennsylvania State College, State College, Pennsylvania
July 15, 1949

THE principle described in the previous communication by W. A. Weyl and D. P. Enright[1] opens new possibilities for producing light sensitive substances. The discovery that asymmetrical force fields are conducive to photolysis led to the synthesis of photosensitive materials by grinding together in a mortar compounds which by themselves were not photosensitive. For example, neither lead oxide nor sodium arsenite darken when exposed to sunlight. A mixture of PbO and Na_3HAsO_3, however, darkens rapidly.

In a similar fashion mercury and bismuth compounds can be made photosensitive. For example, a mixture of Bi_2O_3 with Na_3HAsO_3 or As_2O_3 darkens on exposure to light.

Lead bromide becomes photosensitive when intimately mixed with Al_2O_3, ZrO_2 or ZnO.

These results may explain why the paintings of some artists were found to be more susceptible to darkening than those of others. Pigments such as lead carbonate (white), lead stannate (yellow), lead antimonate (yellow), or mercuric sulfide (red) may become photosensitive when intimately mixed with certain other pigments.

* Sponsored by the ONR and carried out under Contract N6 onr 269–11, NR 081–060.
[1] See previous letter by W. A. Weyl and D. P. Enright, J. Opt. Soc. Am. 39, 889 (1949).

On a Paper by Vašiček concerning Reflection from Multilayer Films

W. L. Wilcock
University Observatory, St. Andrews, Scotland
July 12, 1949

DURING a study of the reflectance of multilayer films using the method given by Caballero[1] I noticed that the results did not agree with those derived by Vašiček[2] from what is essentially a generalization of this method. Subsequent reading showed that Crook[3] has already suggested Vašiček's work to be at fault, but it seems worth while to make clear that his error arises from mishandling the formula for reflection from a single layer, and does not invalidate the reduction of a multilayer film to an optically equivalent single reflecting surface.

Using Vašiček's notation, if a ray of light vibrating in the plane of incidence and of unit amplitude is incident upon a single layer,

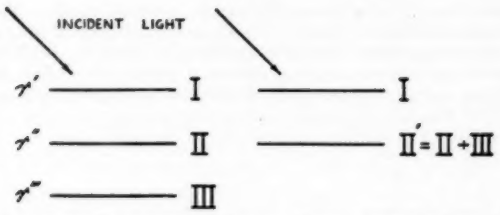

FIG. 1. Reduction of reflection from two layers to reflection from one equivalent layer. (Vašiček's second treatment.)

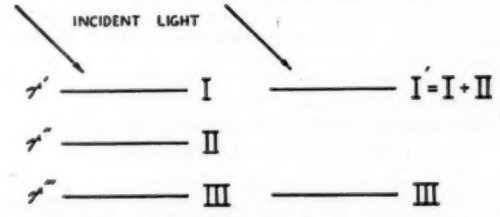

FIG. 2. Reduction of reflection from two layers to reflection from one equivalent layer. (Vašiček's first treatment.)

the resultant reflected ray is given correctly by his Eq. (3) as

$$r_p e^{i\delta_p}=\frac{r_p'+r_p''e^{-iz}}{1+r_p'r_p''e^{-iz}}. \qquad (a)$$

The terms r_p', $r_p''e^{-iz}$ in the numerator are Vašiček's "basic amplitudes" and represent the first reflected rays from each boundary, supposed acting alone; but it is apparent from the derivation of (a) that it is the phase *difference* between these rays which enters into the denominator term $r_p'r_p''e^{-iz}$, and that this here equals the product of the "basic amplitudes" only because the ray r_p' happens to be chosen as the zero of phase. It follows that in Vašiček's second alternative treatment of the three-boundary problem (Fig. 1), the reflection from the equivalent boundary II' should be

$$\hat{r}_p e^{-i\hat{\delta}_p}=\frac{r_p''e^{-iz_1}+r_p'''e^{-i(z_1+z_2)}}{1+r_p''r_p'''e^{-iz_2}},$$

and the resultant reflection from I and II' is

$$r_p e^{i\delta_p}=\frac{r_p'+\hat{r}_p e^{-i\hat{\delta}_p}}{1+r_p'\hat{r}_p e^{-i\hat{\delta}_p}}$$

$$=\frac{r_p'+r_p''e^{-iz_1}+r_p'''e^{-i(z_1+z_2)}+r_p'r_p''r_p'''e^{-iz_2}}{1+r_p'r_p''e^{-iz_1}+r_p'r_p'''e^{-i(z_1+z_2)}+r_p''r_p'''e^{-iz_2}}$$

which agrees with Crook's result.

In his first alternative treatment of the three-boundary problem (Fig. 2), Vašiček's Eq. (12) gives correctly the first reflection from the equivalent boundary I' as

$$r_{1p}e^{i\delta_{1p}}=\frac{r_p'+r_p''e^{-iz_1}}{1+r_p'r_p''e^{-iz_1}}.$$

Reflections between I' and III, however, take place from the opposite side of I', for which the reflection factor can be shown to be

$$r_{1p}'e^{i\delta_{1p}'}=-\left(\frac{r_p'+r_p''e^{iz_1}}{1+r_p'r_p''e^{-iz_1}}\right)\neq -r_{1p}e^{i\delta_{1p}}$$

and because of this inequality the relation (a) cannot be applied to the reflection from I' and III. Using the transmission factors for the two directions across I', whose product is

$$\frac{(1-r_p'^2)(1-r_p''^2)}{(1+r_p'r_p''e^{-iz_1})^2}$$

the reflection from I' and III is found as

$$r_p e^{i\delta_p}=r_{1p}e^{i\delta_{1p}}+\frac{(1-r_p'^2)(1-r_p''^2)}{(1+r_p'r_p''e^{-iz_1})^2}r_p'''e^{-i(z_1+z_2)}$$

$$\times\left\{\begin{matrix}1+r_{1p}'e^{i\delta_{1p}'}r_p'''e^{-i(z_1+z_2)}\\ +\cdots\end{matrix}\right\}$$

$$=r_{1p}e^{i\delta_{1p}}+\frac{(1-r_p'^2)(1-r_p''^2)r_p'''e^{-i(z_1+z_2)}}{(1+r_p'r_p''e^{-iz_1})^2(1-r_{1p}'e^{i\delta_{1p}'}r_p'''e^{-i(z_1+z_2)})}$$

$$=\frac{r_p'+r_p''e^{-iz_1}+r_p'''e^{-i(z_1+z_2)}+r_p'r_p''r_p'''e^{-iz_2}}{1+r_p'r_p''e^{-iz_1}+r_p'r_p'''e^{-i(z_1+z_2)}+r_p''r_p'''e^{-iz_2}}$$

which is the result obtained above.

Both ways of successively reducing the number of boundaries therefore lead to the same result, but it is evident that the treatment is simple only if the substitution of equivalent boundaries is made upwards from the lowest layers.

[1] D. L. Caballero, J. Opt. Soc. Am. **37**, 176 (1947).
[2] A. Vašiček, J. Opt. Soc. Am. **37**, 623 (1947).
[3] A. W. Crook, J. Opt. Soc. Am. **38**, 954 (1948).

Erratum: The Plastic Deformation and Crystal Orientation of Thallium Halides

[J. Opt. Soc. Am. **39**, 445–453 (1949)]

ALEXANDER SMAKULA AND MYRON W. KLEIN
Engineer Research and Development Laboratories, The Engineer Center, Fort Belvoir, Virginia

THE last sentence on page 448 is incomplete and gives an erroneous meaning. The complete sentence should read:

In NaCl and other face centered cubic crystals the glide planes (T) are parallel to the cubic faces and the glide directions (t) parallel to the dodecahedron faces.